TITLE: THE INVESTIGATION OF A NOVEL PROTEINASE INHIBITOR AS A MEANS TO TRANSFER INSECT RESISTANCE TO PLANTS

by

Jacqueline Ruth Tasker

Submitted in fulfillment of the requirements for the Degree of Master of Science in the Faculty of Science, University of Cape Town

Cape Town, November 1995

TITLE : THE INVESTIGATION OF A NOVEL PROTEINASE INHIBITOR AS A MEANS TO TRANSFER INSECT RESISTANCE TO PLANTS

CONTENTS

ACKNOWLEDGEMENTS

I am grateful for the contributions of my supervisors, Professor Jennifer Thomson, Dr. Dennis Maeder and the late Professor Dawie Botes.

I would also like to thank the technical staff, including Di James for help in sequencing, Di de Villiers and Charles Hendrickse, and a number of people of the Microbiology Department of UCT for helpful discussions.

I am grateful to AECI for a Postgraduate Research Fellowship and to the Foundation for Research and Development (FRD) for a Postgraduate Bursary.

ABBREVIATIONS

AA	amino acid
Amp	ampicillin
BAEE	N-alpha-Benzoyl-L-arginine-ethyl-ester
BBI	Bowman-Birk inhibitor
BPTI	Bovine Pancreatic Trypsin inhibitor
CaMV	cauliflower mosaic virus
cccDNA	covalently closed circular plasmid DNA
cDNA	copy DNA
Ci	Curie
CIP	calf intestinal phosphatase
CPG	controlled pore glass
CTPP	C-terminal triphosphate
dATP	adenosine triphosphate
ΔG	change in Gibbs free energy
dH_20	distilled water
DMAP	diethyl amino pyridine
DMTr	dimethoxytrityl
DNA	deoxyribonucleic acid
dNTP	nucleoside triphosphate
ds	double stranded
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EMBL	European Molecular Biology Laboratories
ER	endoplasmic reticulum
EtBr	ethidium bromide
GCG	Genetics Computer Group
GUS	β-glucuronidase
HPLC	high pressure liquid chromatography
IPM	integrated pest management
IPTG	isopropyl β-D-thiogalactopyranoside

LB	Luria Broth
LHS	left hand side
LTR	left terminal repeat
μ	micro
MBN	mung bean nuclease
MCS	multiple cloning site
mRNA	messenger RNA
NOS	nopaline synthase
*npt*II	neomycin phosphotransferase II gene
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PEG	polyethylene glycol
PI	protease inhibitor
PIIF	proteinase inhibitor-inducing factor
Pir	protein information resource database
pK′	negative log of dissociation constant
PNK	polynucleotide kinase
RBS	ribosome binding site
RHS	right hand side
RNA	ribonucleic acid
RTR	right terminal repeat
SDS	sodium dodecyl sulphate
ss	singlestranded
T-DNA	T-strand DNA
T/C	trypsin/chymotrypsin
T/E	trypsin/elastase
T/T	trypsin/trypsin
TAE	tris-acetate-EDTA
TBE	tris-borate EDTA
TE	Tris EDTA
TFA	trifluoroacetic acid
THF	tetrahydrofuran
TIR	translation initiation region
Tm	melting temperature
tRNA	transfer-RNA
UTR	untranslated region
UV	ultraviolet
X-gal	X-glucuronide

TITLE : THE INVESTIGATION OF A PROTEINASE INHIBITOR AS A MEANS TO TRANSFER INSECT RESISTANCE TO PLANTS

CHAPTER ONE

CONTENTS

1.1 BIOLOGICAL CONTROL OF INSECTS

It has been estimated that despite the implementation of extensive control measures crop losses of approximately 13% are caused by insects world-wide (Hilder *et al.*, 1989a). The chemical revolution in agriculture started after World War II, a broad spectrum of crop protection chemicals being introduced into the market starting in the 1950s. During this time, farms in the United States and Europe became highly mechanised and specialised through monoculture, as well as heavily dependent on pesticides, including the new synthetic organochlorine insecticides. Many problems have since been caused by the development of pest resistance to chemical agents.

Resistance among insect pests developed for two main reasons. Firstly, damage was done to beneficial insects, the predators of insect pests, by the broad-spectrum pesticides and secondly, insect pests acquired resistance to the chemicals, and developed into new "superpests" as their natural enemies were eradicated. Other hazards have included the accumulation of chemicals in food chains, particularly the organochlorines and their persistence in the environment, such as in groundwater contamination. Furthermore, situations have occurred in which more and more pesticide needs to be sprayed to reach the same level of control, greatly increasing production costs. By the mid-1950's, 30% of production costs were being spent on organochlorine insecticides in the US. It is due to these concerns, which are expected to escalate in the future, that increasing legislative pressure against pesticide use is being introduced in the western world, and there has been a call both locally in South Africa and internationally for alternative practices that would make agriculture more sustainable (Shields, 1987; Vaeck *et al.*, 1987; Flora, 1990; Reavey and Gaston, 1991). One of these strategies is the biological control of plant pests, to be implemented as an integrated part of pest management (Keetch, 1990a and b; Spurrier, 1990).

Integrated pest management (IPM) is a strategy that was first successfully applied in the mid-1950's in a developing country, Peru, in the Canete Valley, where it is still being successfully used by the major landowners in growing cotton (Sattaur, 1988).

About 8% of US farms now use IPM methods as part of the drive for more sustainable agriculture, and development agencies of first-world countries are promoting and sponsoring a number of projects in developing countries, such as among Asia's small-scale farmers (Madeley, 1994) .

A viable IPM programme involves a clear understanding of and the use of a number of components, the most important being the bionomics of insect pests; monitoring systems to establish the prevalence and seasonal occurrence of insect pests; calculation of economic thresholds; the biology of parasites and predators; utilisation of insect resistant plant varieties; and methods to maximise the advantages of pesticides and minimise their disadvantages. Insect-resistant crop varieties form an important part of IPM. There is usually no extra cost to the farmer once the resistant variety has been obtained and it is easily available to him thereafter. In the context of an IPM strategy, plant resistance should improve the impact on a pest population when both biological and chemical control methods are used.

The transfer of insect tolerance into crop plants is a goal of significant interest to agricultural biotechnology (Brunke and Meeusen, 1991). A logical approach to this goal is through the expression of insect control protein genes from various sources, in transgenic plants. The relatively recent ability to produce transgenic plants by a variety of methods (Bevan, 1984; Hoekema *et al.*, 1985; Hooykaas, 1989) has greatly advanced this goal by enabling the transfer of insect resistance to plants in situations in which conventional methods have failed, in an isogenic setting- other characteristics of the crop remaining unaffected. Conventional plant breeders have historically used germplasm from the wild relatives of cultivated plants to breed resistance into their elite high-yielding crop strains. Today much of the germplasm that plant breeders use to improve crops comes from stocks of their wild relatives found in developing countries. Of the many effective natural plant defence strategies developed over at least a million years of coexistence with insect predators, those most sought after for use in cloning strategies are those depending on a primary gene product for effect and showing simple Mendelian inheritance. One of these strategies includes the production of proteinase inhibitors (PIs). The

utilisation of PIs for transfer of pest resistance to plants leads to a consideration of the context in which PIs interact, and proteolysis itself.

2.1 PROTEOLYSIS

The role of proteolysis in the stability, availability, activation and regulation of proteins in cells is extensive (reviewed by Neurath and Walsh, 1976; Neurath, 1989; Gottesman and Maurizi, 1992; Vierstra, 1993). A wide variety of proteolytic mechanisms enable the cell to respond flexibly to external stresses and developmental stimuli and allow for balanced growth.

Proteolysis is a hydrolytic, and under normal physiological conditions, exergonic and irreversible reaction. It thus effects unidirectional changes in the cellular environment and can induce new cellular functions. New physiological functions it is known to initiate include blood coagulation, fibrinolysis, complement activation, hormone production, metamorphosis, supramolecular assembly and digestion. The need for the control of proteolysis within the cell is obvious and is achieved by a variety of mechanisms. Apart from their roles in digestive proteolysis and protein degradation processes, the regulatory function of proteolytic enzymes through restricted peptide bond cleavage has recently become more evident. Proteolytic processing is an efficient mechanism to regulate physiological processes as it responds rapidly to a by creating an irreversible change that directs the destination (in translocation) or the timing of activation (the appearance of activity) of the target protein. Both of these mechanisms are manifested at the post-translational level. Thus, proteolysis plays an important role in regulation, acting faster than transcription initiation, and triggered by signals operating entirely at a post-translational level.

2.1.1 Functions of proteolytic enzymes

Proteinases are endopeptidases used in biological systems to catalyse the cleavage of specific covalent bonds within proteins. Their specificity enables their action to be confined to a certain subset of peptide bonds at specific sites and in specific

substrates. The four catalytic classes of endopeptidases which are classified according to the residues in their active sites, are the aspartic acid, cysteine, metallo- and serine proteinases (Webb, 1984). (A few other endopeptidases also exist which do not fall under this definition). For each of the above families the groups involved in substrate binding, within the active centre of the enzyme are: two aspartic acid residues, an essential cysteine residue, metal ions, (usually zinc) or a uniquely reactive serine side-chain, respectively (Neurath and Walsh, 1976; Bond and Butler, 1987).

Examples of the aspartic proteinases are cathepsin D and renin. This family includes the lysosomal enzymes and gastric juice enzymes. Among the cysteine proteinases of eukaryotes, the papain superfamily predominates. The family of metalloproteinases includes thermolysin, fibroblast collagenase, signal peptidase from the mitochondria and endoplasmic reticulum, the enkephalinases and gelatinases. The serine proteinases include trypsin, chymotrypsin and elastase. Of these, two major sub-families exist, the trypsin-like and the subtilisin-like enzymes, serving functions in digestion, blood clotting, fibrinolysis, in immune reactions and in fertilisation.

Proteolytic processing occurs as the final step in the biosynthesis of a great variety of proteins, as the directed cleavage of specific peptide bonds in the precusor of the target protein. There is a large variety of physiological processes regulated by such processing. Its physiological significance was first recognised in studies of the activation of zymogens of pancreatic proteases notably trypsin and chymotrypsin. "Limited proteolysis" as it occurs in zymogen reactions is the last step in the synthesis of a number of biologically active proteins, or the first step in protein degradation. Trypsin, chymotrypsin and elastase are produced by the pancreas as zymogens, their inactive precursor forms. These are subsequently converted to their physiologically active forms by the selective enzymatic cleavage of specific peptide bonds. Other examples of zymogens are ribonuclease, and proinsulin, converted to insulin. In general, proteolytic processing releases a conformational restraint on the inactive precursor to generate a biologically active protein. The

processing sites are in relatively flexible interdomain segments or surface loops (Atkinson *et al.*, 1993).

Proteolysis plays an important role in the degradation of abnormal proteins resulting from mutations, biosynthetic errors, disease, or environmentally-caused damage such as from excessive heat or cold. Proteolysis also assists in the stochiometric accumulation of enzymatic subunits and the maintenance of correct enzyme-cofactor ratios. For example, chlorophyll a/b binding protein and plastocyanin are rapidly catabolised in the absence of their respective cofactors, chlorophyll and calcium ions (Mullet *et al.*, 1990).

Proteolysis is also responsible for the removal of signal and transit peptides following the transport of proteins into the endoplasmic reticulum (ER), mitochondria or chloroplasts, and for the generation of peptides that act as hormones, antigens or other effectors. For example, Ryan and coworkers (1990) have shown 200 amino acids are cleaved from the precursor protein prosystemin to yield its active 18 amino acid fragment, systemin, the wound-inducible peptide hormone in tomato. Processing of signal peptides regulates the transport mechanism that directs proteins to their site of action. Processing of pro-proteins is regulated by different enzymes and in different cell compartments. In many cases, pro-protein cleavage is catalysed by serine proteases, for example, a carboxypeptidase B-like enzyme removes the C-terminal residue of pro-insulin, generated by the prior action of a trypsin-like protease. In other cases, dipeptidyl aminoproteases and cysteine or aspartate proteases are involved (Neurath, 1989).

In global regulation proteolysis may be used in conferring a short half-life to an enzyme whose levels can thus be precisely controlled depending on the nutritional or developmental state of the cell or plant. This enables the rapid alteration of substrate flow through a pathway by the attenuation of synthesis of strategic enzymes. This type of regulation can even encompass entire metabolic pathways and events in organelles such as conversion of one form to a more mature form, previously important enzymes being degraded as new organellar functions are assumed. The same control mechanism exists for key regulatory proteins whose

presence or absence can determine a particular developmental programme, including effectors of the cell cycle, developmental regulators, homeotic proteins, transcription factors and protein kinases (Finley and Chau, 1991; Rechsteiner, 1991).

2.1.2 Proteolysis in plants and regulation

Protein turnover has been known to occur in plants for over 60 years, but it is only in the last decade that there has been much light shed on the mechanisms and functional roles of protein degradation in plants. These include essential roles not only in protein recycling, in supplying the amino acids necessary for cellular homeostasis and growth, allowing plants to re-utilise amino acids but also in more specific and directed cases, to enable protein content changes during both development and adaptation to new environmental conditions, which is important for plants as they have a sessile lifestyle.

The best example of proteolytic regulation in plants involves phytochrome regulation, which allows plants to respond developmentally to changing light conditions. Phytochrome is synthesised as Pr, a red-light absorbing form. This form is stable and inactive until activated by light to yield the Prl form, which is far-red light absorbing and subject to rapid proteolysis (Vierstra, 1993).In special cases, carnivorous plants have evolved a highly sophisticated process of trapping, digesting and absorbing insects, which provides an additional extracellular source of amino acids for growth (Juniper, 1989).

One of the most important roles of protein degradation is in cellular housekeeping, for example, in the removal of abnormal proteins (Hershko and Ciechanover, 1992). The calculated half-life of total protein within plant cells is 5-8 days (Davies, 1992). Within the lifespan of a typical plant cell, a substantial proportion of new proteins are synthesized from recycled amino acids, protein breakdown accelerating during nutrient deprivation. During leaf and flower senescence, up to 70% of protein may be mobilised in the timed disintegration of specific cells- programmed cell death or apoptosis. It is also used in more restricted ways in leaf abscission, xylem and sclerenchyma differentiation, tapetal disintegration prior to pollen

dehiscence, and testa and periderm maturation (Vierstra, 1993). In mammals, starvation is known to activate the largely indiscriminate transport of proteins into the lysosome, which contains a variety of proteases. The same is thought to happen in plant and yeast vacuoles, regarded as the functional equivalent of the lysosome (Matile, 1982; Yoo and Chrispeels, 1990).

Several types of proteolytic enzymes are involved in the degradation of proteins in germinating seeds. The degradation of seed storage proteins (SSPs) produces much of the amino acid supply necessary for growth of the emerging seedling. The SSPs are synthesised during seed maturation and deposited in protein bodies, a specialised form of the vacuole (Yoo and Chrispeels, 1990). Following seed germination, SSPs are degraded in part by specific proteases synthesised *de novo* and transported into the protein body. Several of the proteases have been isolated and their corresponding genes cloned. Some are related to the thiol proteases, cathepsins, present in mammalian lysosomes. All four types (cysteine, serine, aspartyl- and metallo-) have been purified from germinating seeds, but at germination there is a marked increase in the activity of cysteine proteases in particular, which are responsible for the catabolism of the majority of seed reserve proteins (Shutov and Vaintraib, 1987; Ryan, 1991).

In legumes, *de novo* synthesis of cysteine proteinases during germination has long been recognised although little was known of the mechanisms controlling their regulation. This area has come under intensive investigation in recent years. Cervantes *et al.* (1994) recently identified a factor present in the embryonic axis of chickpea seeds responsible for the control of mobilisation of SSPs. This factor is able to activate *cac* (chickpea cysteine proteinase) expression via ethylene production. Decreasing levels of ethylene are paralleled by decreasing levels of the transcript, which sheds light on the molecular events ethylene may trigger in the control of proteolytic activity in the seed during germination.

Plant cells, like bacteria, often recognise foreign proteins and degrade them, thus preventing adequate expression of heterologous genes after modification of the cell by genetic engineering (Enfors, 1992). As this has important ramifications in crop

improvement attempts via genetic engineering, new emphasis has been placed on understanding the factors that govern protein stability in plants and on identifying the interfering proteases. This problem has been overcome in *Escherichia coli* with the development of *lon*$^-$ strains. These are deficient in the protease (La), which is responsible for degrading most abnormal proteins in this organism (Goldberg, 1992; Maurizi, 1992). Comparable mutants have yet to be created in plants.

1.3 PROTEINASE INHIBITORS

Proteinase inhibitors have the property of forming reversible stochiometric protein-protein complexes with various classes of proteolytic enzymes, effecting competitive inhibition of their catalytic function (Bode and Huber, 1992; Travis and Salvesen, 1983). They are found ubiquitously, being present in many forms in a number of plant and animal tissues and microorganisms.

1.3.1 Plant proteinase inhibitors

Plant proteinase inhibitors have come to be grouped among the defensive proteins in plant tissues that contribute to resistance to insect predation and in some cases, microbial pathogens (Richardson, 1977, Broadway *et al.*, 1986; reviewed by Ryan, 1989, 1990; Geoffroy *et al.*, 1990). The fact that some plant PIs inhibit proteolytic enzymes from a wide variety of microorganisms and insects while only rarely inhibiting proteinases from plants, indicated early on that they may serve a role in the defence strategy of plants against microbial pathogens or insects. Green and Ryan (1972) first demonstrated that serine PIs were inducible in tomato seedlings in response to leaf injury. Walker-Simmons and Ryan (1977a) reported on the wound-induced accumulation of trypsin inhibitor activities in plant leaves from a survey of several plant genera. Graham *et al.*, (1985) reported the induction of tomato proteinase inhibitors, as a result of larval feeding, is invoked as a factor that systemically reduces leaf quality. Many other reports of plant PIs have confirmed these findings (Sanchez-Serrano *et al.*, 1987, Davis *et al.*, 1993).

The occurrence of PIs in plants has been known since 1938 and they are widely found in the plant kingdom, being most studied in the the Leguminoseae, Graminae and Solanaceae (reviewed by Ryan, 1990; Table I).

TABLE I. Protease inhibitor families in plant tissues (Ryan, 1990)

1. Soybean trypsin inhibitor (Kunitz)
2. Bowman-Birk inhibitor
3. Barley trypsin inhibitor
4. Potato Inhibitor I
5. Potato inhibitor II
6. Squash inhibitor
7. Ragi 1-2/Maize bifunctional inhibitor
8. Carboxypeptidase A, B inhibitor
9. Cysteine proteinase inhibitor (cystatins)
10. Aspartyl proteinase inhibitor

The serine inhibitors

The serine inhibitors, having the most potential as plant protective agents due to the class of enzymes they inhibit, are also the most studied group of proteinase inhibitors. They have been classified into at least 11 different families based on sequence similarity, topological similarity, and mechanism of binding (Ryan, 1989; Table II). The Bowman-Birk, Kunitz and Kazal inhibitors are named after their discoverers. In the words of Ryan, "Nature has apparently invented serine inhibitors several times", resulting in at least 13 gene families, classified on the basis of primary structure by homology studies from different families.

TABLE II Families of inhibitors of serine proteinases (Ryan, 1989)

ANIMALS

1. Bovine pancreatic trypsin inhibitor (BPTI; Kunitz)
2. Pancreatic secretory trypsin inhibitor (Kazal)
3. Ascaris inhibitor
4. Chelonianin
5. Serpin (mechanistically distinct)
6. Hirudin

PLANTS

7. Soybean trypsin inhibitor (STI) (Kunitz)
8. Soybean proteinase inhibitor (Bowman-Birk)
9. Potato I (PI-I)
10. Potato II (PI-II)
11. Barley trypsin inhibitor
12. Squash inhibitor

MICROBIAL

13. Streptomyces subtilisin inhibitor (SSI)

The serine PIs of plants are abundant in both dicots and monocots. In these plants they are found in storage organs such as tubers and seeds and/or are induced in leaves upon wounding or insect feeding. As well as their role as protective agents in the inhibition of insect and microbial proteinases, they play a role as regulatory agents, in the control of endogenous proteolysis of storage proteins during seed maturation. Their additional role as sulphur storage proteins has been inferred from, firstly, their presence in large quantities in seeds and tubers, in which they represent up to 6% of soluble proteins of soybean and up to 10% of that of barley grains and tubers (Ryan, 1973). Secondly, they contain a high proportion of cysteine amino acids (approximately 20 percent; Ryan, 1977), which is significant as sulphur amino acids are limiting in plants. Further evidence supporting their storage function is that certain PIs such as the Bowman-Birk inhibitor of mung bean

(Wilson and Tan-Wilson, 1987) and the Kunitz trypsin inhibitor of soybean (Wilson *et al.*, 1988) have been shown to be degraded during germination and seedling growth. Degradation of PIs has been found to occur at susceptible interdomain junctions (Vogel, 1988). The seeds of many plants belonging to the legume family have been found to be rich sources of serine proteinase inhibitors, and these are particularly of the Bowman-Birk type (Laskowski and Kato, 1980).

The Bowman-Birk inhibitors

The Bowman-Birk inhibitors (BBIs) are small peptides of about 80 amino acids, of great interest in their potential as plant protection agents and one of the best characterised of the distinct plant PI families. BBIs have been isolated from the legumes soybean, chickpea, garden bean, lima bean, peanut, garden pea and the wild pea *Macrotyloma axillare* (Flecker *et al.*, 1987; Joubert *et al.*, 1979). They show extensive homology as a family. The BBI's are proposed to have arisen by duplication of a single-headed gene followed by divergence to these three different classes, as the active centre serines are bounded by highly variable amino and carboxyterminal regions. They are called double-headed as they have independent reactive sites located in one of the two di-sulphide bridge-bonded loops, each of which may inhibit the same or different classes of enzymes. They may thus be designated trypsin/trypsin (T/T); trypsin/ chymotrypsin (T/C) or trypsin/elastase (T/E) inhibitors.

Evolutionary history of proteinase inhibitor families

The known plant PIs are not all homologous and the several families of which they consist are proposed to have arisen by convergent rather than divergent evolution (Laskowski and Kato, 1980). They characteristically possess several homologous reactive sites on the same polypeptide chain, which are thought to have been formed by repeated duplication. Their specificity can be altered by the replacement of their reactive site residue (P_1) by another residue. For example, using site-directed and cassette mutagenesis, replacement of Met 73 in the active site of *Streptomyces* subtilisin inhibitor with Lys or Arg results in trypsin inhibitory activity and by Tyr or Trp, in chymotrypsin inhibitory activity (Koijima *et al.*, 1990). In each family the P_1 residue changes frequently, resulting in such changes in inhibitory specificity. In

(Gatehouse and Boulter, 1983; Gatehouse *et al.*, 1984; Burgess *et al.*, 1994). This was found to be due to the overproduction of trypsin, which, together with the insufficient availability of sulphur-containing amino acids needed for enzyme synthesis, was proposed to result in the growth inhibition. There is no evidence of toxicity to humans or other animals, thought to be due to the different organisation of mammalian and insect digestive tracts (Burgess *et al.*, 1994). Studies on the effects of PIs on growth and development of insects either in artificially-defined diets or naturally in plant tissues have shown serine PIs can be detrimental to the growth and development of insects of several genera, including *Heliothis*, *Spodoptera*, *Callosobruchus*, *Diabrotica* and *Tribolium* (Gatehouse and Boulter, 1983; Ryan, 1989).

1.3.3 The use of a Bowman-Birk inhibitor to engineer insect resistance in plants

The viability of using a plant gene to engineer insect resistance has been demonstrated by the use of the cowpea trypsin inhibitor (CPTI) gene by Hilder and coworkers (1987; Hilder *et al.*, 1989b). It was West African farmers who had first noted the superior resistance of a certain variety of cowpea to its major storage pest, *Callosobruchus maculatus*, the bruchid beetle. Gatehouse and coworkers (1979), tested seeds of a number of cowpea varieties (over 5000 accessions) for the presence of various antimetabolic secondary compounds, including lectins, various proteinase inhibitors, and phytoalexins. In the resistant variety of cowpea , Tvu2027, resistance was found to be associated with elevated levels of protease inhibitors in the seeds of this line.

Only inhibition of trypsin and to a lesser extent, chymotrypsin, could be detected. The albumin fraction of the seed was toxic to this pest at a level of 10%, which was equivalent to 0.85% of the purified inhibitor. The albumin fraction lost this activity when the inhibitor was purified out by affinity chromatography. Further investigation revealed a small family of at least four iso-inhibitors was present in this line, T/T inhibitors predominating, although a T/C inhibitor was also present (Gatehouse *et al.*, 1980). The T/T inhibitors were found to have a greater inhibitory effect *in vitro* against lepidopteran gut proteases. They identified as the resistance

factor the antimetabolic protein designated Cowpea trypsin inhibitor (CPTI), a member of the BBI family.

Hilder and coworkers (1987), isolated a cDNA clone of CPTI from a cowpea cotyledon cDNA library. The coding sequence for the precursor was placed under the control of the strong constitutive CaMV 35S promoter and transferred to tobacco. The transgenic tobacco was found to have significant resistance to the tobacco budworm, *Heliothis virescens*.

The CPTI accumulated to physiological levels of 0.5% to 0.9% of total soluble proteins in leaves. Both Coleopteran and Lepidopteran insects were susceptible, in feeding trials. Insects on CPTI-expressing transgenic plants did limited damage to the foliage, and unlike those on control plants, showed reduced growth and mortality (Hilder *et al.*, 1989a). Incorporation of the CPTI gene into other crops of agronomic importance such as maize and rice is feasible (Hilder *et al.*, 1987). The resistance of cowpea seeds to bruchid beetles has been found not to be related to levels of cysteine PIs (Fernandes *et al.*, 1993).

1.4 THE KY ANTITRYPTIC PEPTIDE FROM A BBI OF *MACROTYLOMA AXILLARE*

1.4.1 Design of the antitryptic peptide KY.

In 1979, in South Africa, two proteinase inhibitors were isolated from seeds of an indigenous wild pea plant, *Macrotyloma axillare* (Joubert *et al.*, 1979). The inhibitors, designated D3b and DE4 were found upon amino acid sequencing to be homologous with other known members of the Bowman-Birk family, having extensive homology with members of this family (Fig. 1.1) and sharing inhibitory specificities, invariant residues and independent reactive sites (Fig 1.2).

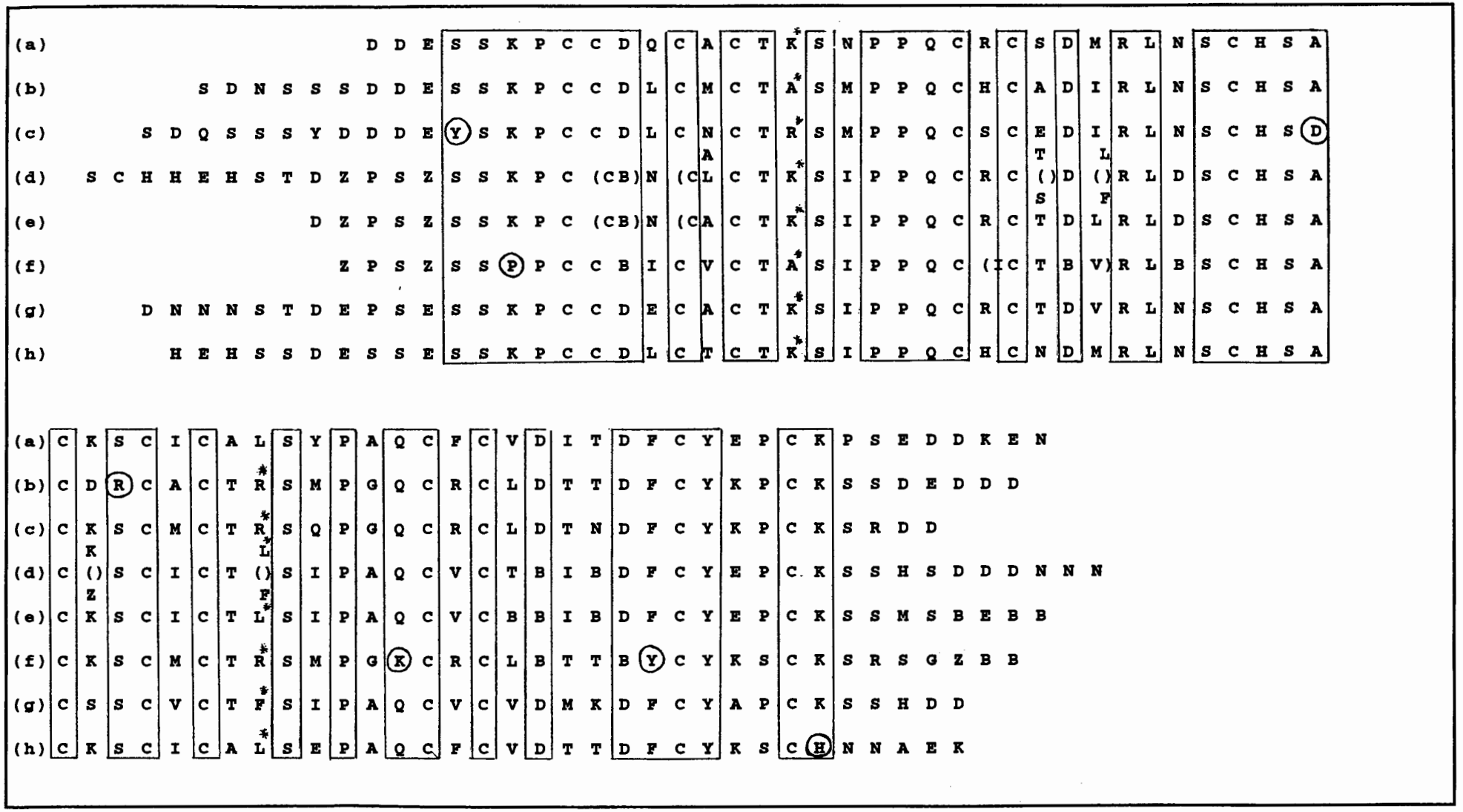


Fig. 1 .1 Comparison of the amino acid sequences of the inhibitory reactive sites of various Bowman-Birk inhibitors (Joubert *et al.* 1979) a, Soybean BBI; b, Soybean CII; c, Soybean DII; d, Lima Bean OI1-IV; e, Lima Bean OIA-I; f, Garden Bean II and II; g, *M. Axillare* seed DE-3; h, *M.axillare* seed DE-4 The positions of invariant amino acids are boxed. The circles in boxed regions indicate variant residues. The reacive site residues are shown by asterisks.

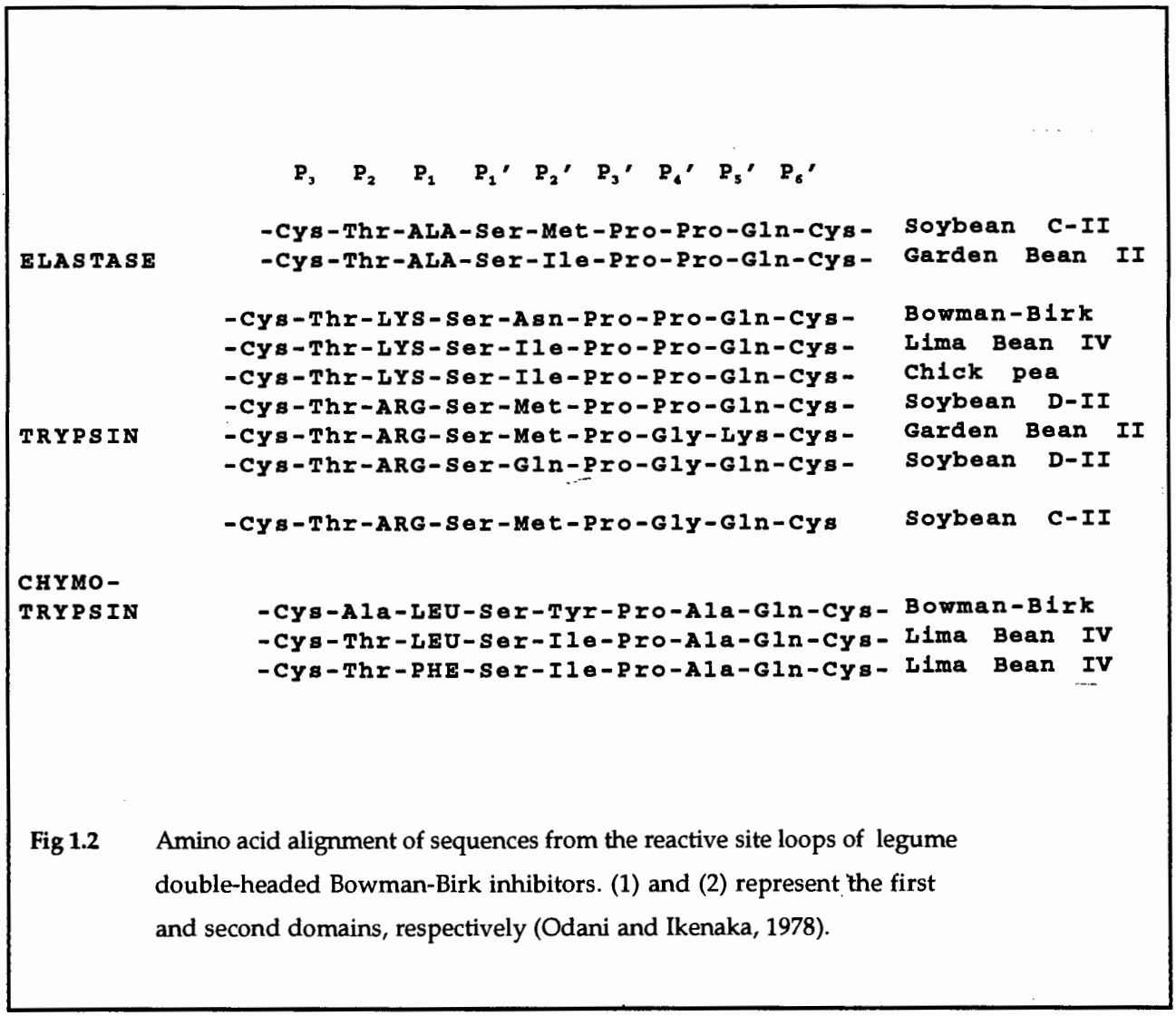


Fig 1.2 Amino acid alignment of sequences from the reactive site loops of legume double-headed Bowman-Birk inhibitors. (1) and (2) represent the first and second domains, respectively (Odani and Ikenaka, 1978).

The antitryptic activity of the double-headed inhibitor DE4, of 76 amino acid residues, was found to be vested within an outer nonapeptide loop. Cleavage of DE4 with cyanogen bromide and pepsin yielded two active fragments with independent inhibitory activity against trypsin and chymotrypsin (Fig. 1.3; Townshend *et al.*, 1982). The antitryptic loop was found to contain the Lys24-Ser25 reactive bond, identified as the P_1 and P_1' residues characteristic of this group of inhibitors, those of the antichymotryptic loop being Leu50-Ser51.

The antitryptic loop was modelled to the active site of trypsin to deduce possible changes which would increase its catalytic efficiency. This enabled the design of a more potent inhibitor.

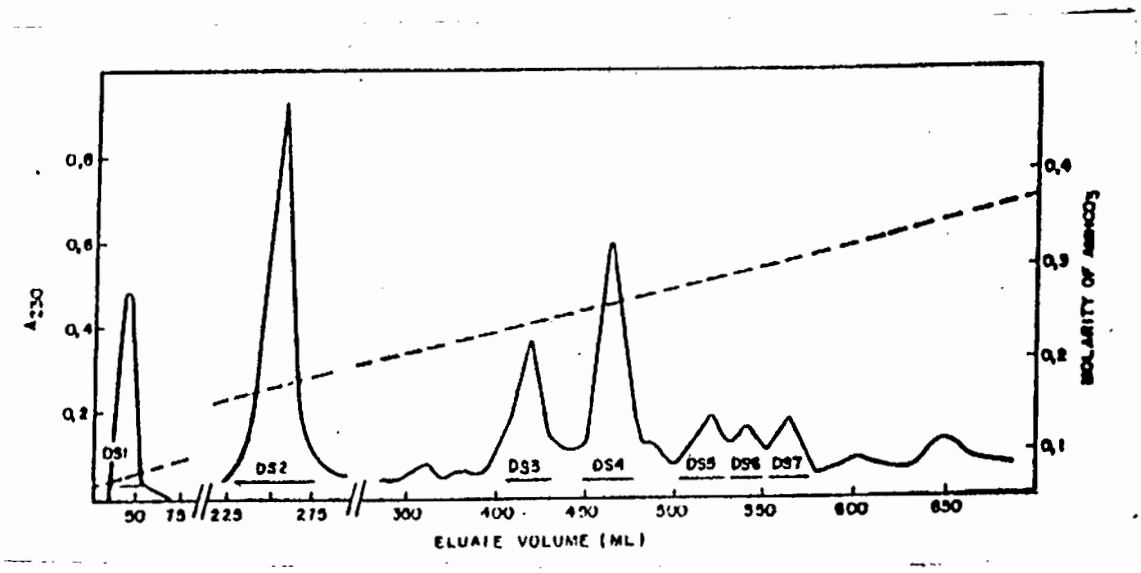


Fig. 1.3 Diagram representing the cleavage of D4 into two separate inhibitory fragments. F-t and F-c inhibit trypsin and chymotrypsin, respectively. The highlighted section represents the 13-mer antitryptic loop (CTCTKSIPPQCHC) (Townshend *et al.*, 1982).

If the small antitryptic loop of DE4 were found to have inhibitory activity in its own right, it was reasoned that this portion of the molecule could be sufficient to confer increased protection against pests when expressed in plants (Sunde, 1989). It was considered that the cost and complexity of the production of a synthetic inhibitor could be reduced and the process of incorporating the coding sequence into plants would be simplified. A model for the interaction of the Bowman-Birk antitryptic loop with trypsin during complex formation was developed to elucidate the stereochemical reasons for inhibitory potency and specificity (Maeder *et al.*, 1992). The model also served as a valuable tool for the design of possible improved analogues with increased activity.

A series of peptides were modelled based on the sequence of the DE4 anti-tryptic loop, residues 19-31 (STCTKSIPPQCHS). The rationale for particular amino acid replacements was based on energy minimisation dynamics and computer-assisted modelling. Modelling was done by homology based on information from the

Brookhaven Protein Databank with an Evans and Sutherland PS390 molecular graphics system, using Biograf PS300 version 6.6 software to produce three-dimensional structures. As no example of a crystallographically determined three-dimensional structure for a Bowman-Birk inhibitor had been lodged with the Brookhaven Protein database, the structure of the binding loop was modelled on the structures of known serine proteinase inhibitors and their complexes. These were, for the inhibitor, alpha-1 proteinase inhibitor (alpha1PI), two domains of the turkey ovomucoid inhibitor (Kazal family) and for the formation of the trypsin-loop complex, the bovine pancreatic trypsin inhibitor (BPTI) and bovine trypsin complex, and the complex of porcine trypsin with soybean trypsin inhibitor (Sunde, 1989).

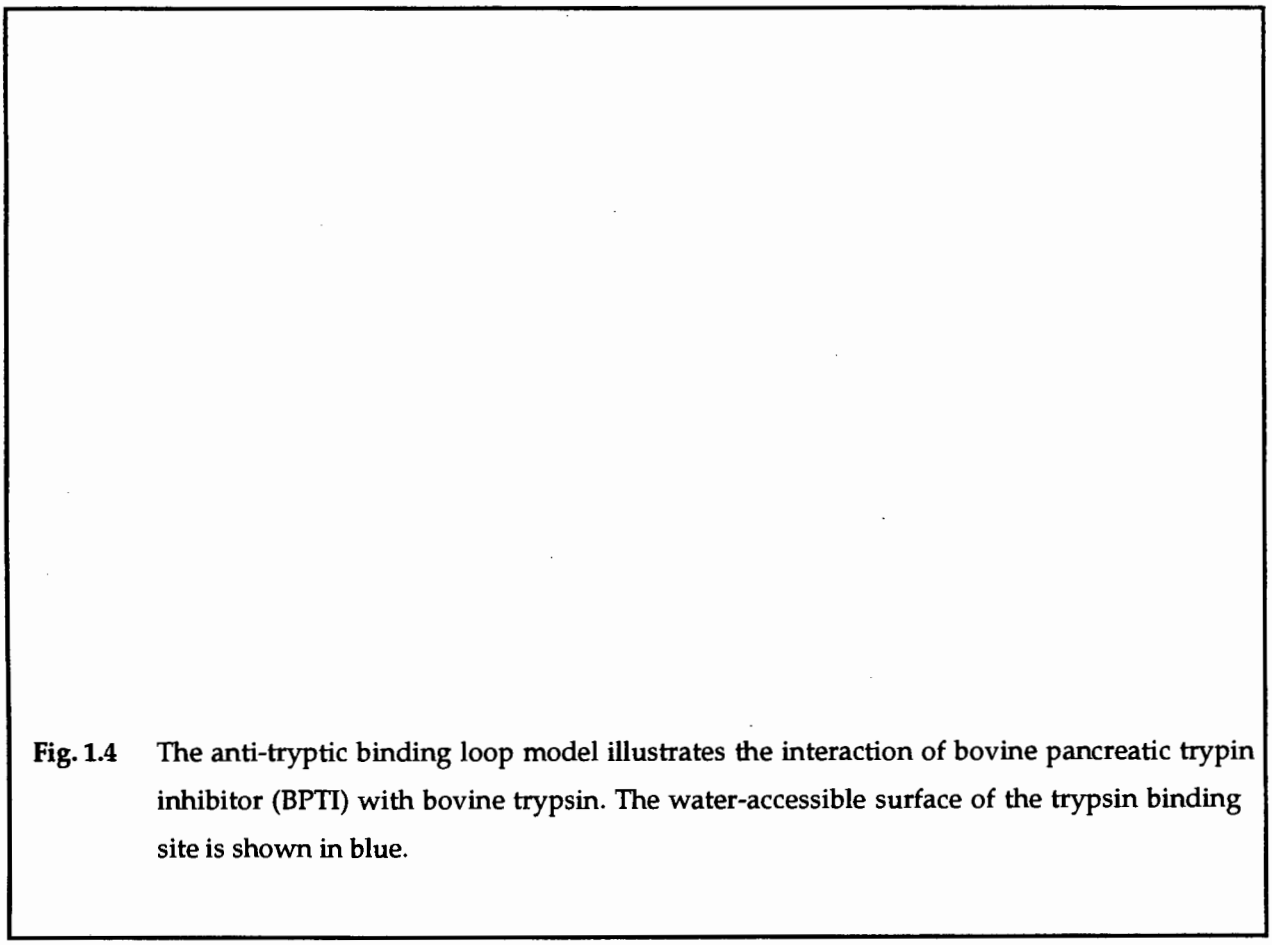


Fig. 1.4 The anti-tryptic binding loop model illustrates the interaction of bovine pancreatic trypin inhibitor (BPTI) with bovine trypsin. The water-accessible surface of the trypsin binding site is shown in blue.

The modelled form of the enzyme-inhibitor complex for trypsin with the DE4 loop indicated the mechanism of inhibition and suggested structural reasons for inhibitory specificity and potency. The model indicated a steric hindrance in the P_5 -

P_4 region flanking the antitryptic loop. Two analogues (variants of the original loop), with sequence changes introduced to minimise this clash, were synthesised to test the validity of the predictions. These were ATL ("Atloop") and KY, (so-named after the K and Y amino acid replacements of this analogue). In a parallel study, a model of the three-dimensional structure of the antitryptic loop was constructed in a knowledge-based manner, by comparison with homologous proteins whose structures were known from X-ray diffraction analyses.

A 13 amino acid peptide corresponding to the amino acid sequence of the inhibitor was first synthesised. It was identical to the active site except that the two terminal cysteine residues were changed to serine to prevent possible incorrect disulphide bridges from forming. As Ser is stereochemically equivalent to Cys, but does not have the disadvantage of disulphide bonding, this enabled the retention of just one disulphide bond. This substitution was made so that disulphide bridges formed only between residues 1-16 and 6-14 of the synthetic peptide. This peptide was shown to have potent anti-trypsin activity *in vitro*. The sequence (ATL, or Atloop), was:

1	2	3	4	5	6	7	8	9	10	11	12	13
Ser	Thr	Cys	Thr	Lys	Ser	Ile	Pro	Pro	Gln	Cys	His	Ser
S	T	C	T	K	S	I	P	P	Q	C	H	S

For the second analogue, KY, an improvement was indicated to be be possible if the bulky beta-branched Thr at residue 2 were removed. This aimed to minimise steric clashes and enable a closer fit to trypsin. In addition, the last two amino acids His and Ser were replaced with a Tyr residue as they were considered redundant, at the same time providing a spectroscopic marker at the carboxy terminus. This enabled subsequent quantitative kinetics determinations in *in vitro* experiments with synthesised peptide. These changes yielded the 11 amino acid sequence analogue KY:

1 2 3 4 5 6 7 8 9 10 11

Ser-Cys-Thr-Lys-Ser-Ile-Pro-Pro-Gln-Cys-Tyr

S C T K S I P P Q C Y (KY)

These alterations can be seen in the context of the labelled residues below, the disulphide bridge pair (3 and 6'), the active site (1 and 1'), residues within the loop and bound to trypsin (2 and 2'), and the carboxy terminus, 7', the dashed line indicating the position of the disulphide bridge.

|----------|

S T C T K S I P P Q C H S (ATL)

S C T K S I P P Q C H S (KY)

5 4 3 2 1 1' 2' 3' 4' 5' 6' 7' 8' (P residue)

The standard method for assaying proteinase inhibitors utilises spectrophotometry to determine the inhibition of the enzymatic hydrolysis of natural or synthetic substrates. Anti-enzymatic activity is assayed after preincubation of enzyme with inhibitor. For the KY inhibitor assays, Ki values were determined from the initial rates of hydrolysis by porcine pancreatic trypsin with its substrate, N-alpha-Benzoyl-L-Arginine ethyl ester (BAEE). The change in release of substrate, detected as an increase in A_{253} was compared in the absence and the presence of the inhibitor. The synthetic KY analogue was tested and found to have an inhibition constant of 1×10^{-7} compared with 1.03×10^{-6} of Atloop. Linear forms of both peptide inhibitors showed no competitive inhibition (Maeder *et al.*, 1992). The K_i for the original peptide ATL was 0.93 μM +/- 0.57 μM. K_i determinations from studies comparing the analogues (Sunde, 1989) were: ATL, 1.03 μM; KY, 0.10 μM and RY, 1.32 μM . Thus, the efficiency of inhibition of KY was improved by an order of magnitude over that of ATL.

1.4.2 Considerations for the design of KY sequence constructs for expression in *Escherichia coli*

The features of heterologous constructs necessary and beneficial for expression in *Escherichia coli* are well established. Many studies have been done and reviews written on prokaryotic expression in this model host bacterium (Stormo *et al.*, 1982; Carrier 1983; Glick and Whitney, 1987; Das, 1990; Stader and Silhavy, 1990; Schatz and Beckwith, 1990) and expression vectors developed (Friesen and An, 1990).

A specific ribosome binding site just prior to the translation start codon is required for the initiation of translation of mRNA in *E. coli*. The ribosome binding site or Shine-Dalgarno (SD) sequence is complementary to the sequence (5'-ACCUCCU- 3') at the 3' end of the 16S rRNA, and thus plays an important role in guiding the small 30S subunit of the ribosome to the initiator codon. The 13 nucleotides at the 3' end of 16S rRNA are: 5'-GAUCACCUCCUUA-3'. These are thought to be unpaired and available for annealing to the mRNA. An extensive comparison with 124 *E. coli* translation initiation site sequences showed that all but one had, within the 15 bases preceding the initiation codon, a complementarity of at least 3 bases to the 16S 3' end (Stormo *et al.*, 1982). The average length of complementarity was 5+/-1 bases. The spacing between the complementary bases and the initiation codon was 7+/-2 bases. Of the 13 bases of the 3' end, the CCUCCU subsequence was found to most often complement the mRNA sequence. AUG has been found to be the most efficient codon for the translation start signal (Stormo, 1986; Das, 1990).

The SD sequence in combination with the start codon forms the translation initiation region, (TIR), which also includes 20 nucleotides up and downstream of these two regions. Over this region, the sequence itself is not a determining factor in translational efficiency, but rather accessibility of the transcript to the ribosomal machinery due to the lack of secondary structure. There must also be a stretch of between 10 to 18 nucleotides before the SD region, or no translation initiation will occur.

In *Escherichia coli* the consensus sequence for the promoter region consists of a -35 box and a -10 box, with the following general format.

5' .....TTGACA.....TATAAT.....3'

-35 -10

The closer the promoter to this sequence the stronger it is in its ability to direct transcription.

While theoretically stable mRNA structures do not necessarily cause translation delays (Sorenson *et al.*, 1989), some workers have found changes such as the addition of a leader sequence upstream of the gene of interest to be necessary to overcome negative effects of tightly folded mRNAs during translation (Schoner *et al.*, 1984; Sung *et al.*, 1986). Site-directed mutagenesis of 5' untranslated regions (UTRs) has also been successfully used to enhance translational initiation by destabilising stem-loop structures. Another strategy is the inclusion of an upstream cistron with an extra RBS that will help to produce a virtually unstructured mRNA 5' end to facilitate the accessibility of the ribosome to the RBS.

Systematic codon preferences have been widely reported for bacterial, yeast, plant and mammalian genes, as well as the trend of highly expressed genes correlating with increased codon bias compared to codons of weakly expressed genes (Mullet, 1989; Anderson and Kurland, 1990). Studies of *E. coli* have supplied the strongest evidence of codon bias, the major trend being a strong bias towards a particular subset of codons in highly expressed genes to more even codon usage in genes expressed in low amounts (Sharp *et al.*, 1988). In genes within a taxonomic group there exists a species-specific 'codon dialect' and within genes of a single species, bias in codon choice appears to be related to the level of expression of the protein encoded by that gene. Codon usage in highly expressed genes has been found to correlate with the abundance of isoaccepting tRNAs.

Codon bias is a consideration when engineering expression of foreign genes in heterologous systems. It has been proposed that the cell uses rare codons to differentiate regulatory and nonregulatory genes. Rare codon usage may be part of a

general mechanism for modulating protein products that cannot be tolerated in the cell in excess amounts, while expression of the adjacent genes on the same mRNA transcript are high (Konigsberg and Godson, 1983). Makoff and coworkers (1989) found a fourfold improvement in expression of Tetanus toxin fragment C in *E. coli* on removing rare codons. However, there are many examples of high level expression in *E. coli* of heterologous genes with an unfavourable codon bias and it may be the case that rare codons are more likely to limit expression if they are clustered (Makoff *et al.*, 1989). Considering that transcription and translation are coupled in prokaryotes, and as rare codons are known to be translated more slowly, pausing of the ribosome in these areas of rare codon frequency are proposed to expose stretches of the mRNA, making it susceptible to endoribonucleases, and rho-dependent termination. This may explain the apparent stabilisation of the transcript on removing rare codon sequences in bacteria.

All secretory proteins of prokaryotes and eukaryotes, including integral membrane proteins, periplasmic and secreted proteins have a signal sequence, whose primary function is to help direct them into or across the cytoplasmic membrane. Classic hypotheses as to how this is achieved (Blobel and Dobberstein, 1975; Von Heijne and Blomberg, 1979) are being further revised (as reviewed by Pugsley, 1993). The similarities between signal sequences of different proteins are in their general structural features and not their specific sequence. The N-terminal domain is a short, generally hydrophilic domain with one or two positively charged residues thought to be important in enabling insertion of the signal, by accomodating electrostatic interactions with negatively charged phospholipid head groups of the membrane. This region is followed by a longer hydrophobic domain, a stretch of 5-10 highly hydrophobic residues near the middle of the signal peptide. The hydrophobic core region, also essential for function (Talmadge and Gilbert, 1989).

The cytoplasm of *E. coli* is too reducing to allow the formation of correct disulphide bonds. Partial disulphide bond formation is known to occur only within aggregates in the cytoplasm, attributed to the formation of oxidising conditions within the aggregates (such as in the production of calf-prochymosin in *E. coli* (Holmgren *et al.*, 1979). In these studies, pulse-chase and PAGE analysis indicated oxidation of

sulphydryl groups is not possible in the cytoplasm. Secretion is therefore of vital importance to this process in allowing translocation of disulphide bond-rich proteins from the reducing cytoplasmic environment into the more oxidised periplasmic space. *In vitro* studies have shown that the thioredoxin system of *E. coli* is able to reduce and oxidise protein disulphides reversibly. The first step appears to be random formation of disulphide bonds and formation of a loosely folded structure followed by disulphide interchange catalyzed by thioredoxin facilitating formation of the correct disulphides. NADPH is involved as a reducing agent in this process (Holmgren, 1979).

There are several advantages of periplasmic localisation of the peptide in this study. The periplasm offers the appropriate redox environment for disulphide bridge formation. This may enable folding and hence activity of the protein (Stader and Silhavy, 1990). Foreign proteins are known often to be unable to attain their native conformation in the cytoplasm of *E. coli*, especially where the structure is stabilised by disulphide bridges (Goeddel, 1990), as these cannot form in the reducing conditions found here. There is an oxidising environment in the periplasm conducive to correct disulphide bridge formation. Loop formation by disulphide bridging is essential for the inhibitory activity of the synthetic fragment KY and protects it from hydrolysis. Linear forms of both peptide inhibitors modelled from DE4 (Joubert *et al.*, 1979) showed no competitive inhibition (Maeder *et al.*, 1992). Export to the periplasm may protect foreign proteins which are unstable due to proteolytic degradation. Proteolytic enzymes are unevenly distributed in the cell compartments. In total eight soluble endogenous proteases have been characterised in *E. coli*. Two are periplasmic, five are found in the cytoplasm and one is equally distributed between the two compartments (Swamy and Goldberg, 1985).

Small recombinant peptides, which are more susceptible to proteolytic degradation in the cytoplasm are stabilised by attachment of signal sequences for periplasmic targeting (Derynk *et al.*, 1984). Molecules transported to the periplasmic space are at least 10 times more stable than those that remain in the cytoplasm (Talmadge and Gilbert, 1982). Periplasmic expression has the additional benefit of releasing the initiator f-Met from the end of the protein. This is not always correctly removed

from recombinant proteins expressed in *E. coli*. Thus, the N-terminus of the recombinant protein would be that of the natural protein, provided that the signal peptidase cleaves between the signal sequence and the natural N-terminal amino acid. As periplasmic proteins normally account for only 4% of total cell protein, this strategy simplifies purification procedures.

Signal peptides are processed during translocation across the cytoplasmic membrane by cleavage at a defined recognition site by signal peptidases. A cytoplasmic membrane protein of *E. coli*, Protease IV, has been shown to act as a signal peptidase, specifically cleaving the signal peptide from the mature protein during passage through the inner cell membrane (Ichihara *et al.*, 1984). Host cell factors are known to be important in maintaining the translocation-competent state of secretory proteins. Although there is undisputed evidence that signal sequences are an absolute requirement for the export of such polypeptides, it has also been found in practice that a signal sequence alone does not suffice for translocation of a protein across the plasma membrane (Michaelis and Beckwith, 1982; Silhavy *et al.*, 1983; Kadanoga *et al.*, 1984). The charge of residues following the signal sequence is a potentially important feature. Summers (1988) found t-isomerase from chicken could not be secreted or processed when fused to the β-lactamase signal peptide unless the Arg at position 3 of the mature protein was replaced. This was proposed to have been due to the extreme pK' of Arg. This problem was overcome by ion-pairing, by insertion of 2 Glu residues in front of Arg at the signal peptide junction.

A strategy used to overcome the problem of misfolding is the cotranslational secretion of the protein to an oxidising environment, either the periplasm or the extracellular medium. Although this is successful for some proteins, for others it leads to cell death (Kohno *et al.*, 1990). A second strategy is the production of the protein in the cell in an unfolded state followed by subsequent *in vitro* folding. This can be by utilising air oxidation, catalysed by the presence of trace metals, disulphide isomerase, or a mixture of reduced and oxidised thiol compounds such as oxidised and reduced DTT or glutathione (GSSG and GSH respectively). The recovery of artificial PI constructs produced in *E. coli* has been achieved by the purification of unfolded protein from the cytoplasm and subsequent folding by

controlled oxidation. Examples include the production of *Cucurbita maxima* trypsin inhibitor, CMTI from squash (Kupryszewski *et al.*, 1986), the prototype Bowman-Birk inhibitor from legumes (Flecker *et al.*, 1987), and the protease inhibitor eglin c from the leech, *Hirudo medicinalis* (Rink *et al.*, 1984). This method would not seem to be viable for the KY inhibitor. While the folded KY peptide would be recoverable by affinity chromatography on a trypsin column, it would not be possible to recover the small peptide from the cytoplasm in an unfolded state.

1.4.3 Considerations for the design of KY sequence constructs for expression in plants

The successful expression of a foreign gene in plants can be predicted to be the result of many different factors, including an efficient vector, the appropriate promoter, the leader sequence, the 3' noncoding sequence, potential 'volunteer' plant regulatory sequences, codon frequency, secondary structure of the mRNA, and the gene product itself. Many of these factors have influenced the design of vectors used to obtain the high levels of expression of a number of proteins in plants (Bevan, 1984; Deblaere *et al.*, 1987; Mitage *et al.*, 1988; McBride and Summerfelt, 1990). For mRNA coding genes of eukaryotes, which are transcribed by RNA polymerase II, the promoter region may be divided into several components: the cap site element located around nucleotide -1 (the start site of transcription), the TATA box located around -30, the upstream elements between about -40 and -110, and enhancer elements that are located either further upstream or downstream from the initiation site.

The TATA box, a positional element that directs the transcription machinery to initiate RNA synthesis from a given promoter region consists of a highly conserved AT-rich region with the consensus sequence 5' TATAA/TAA/T-3', located approximately 30 bp upstream of the start of transcription of most eukaryotic protein coding genes (Brethnach and Chambon, 1981). The upstream promoter elements -40 to -110, 5' to the TATA box, determine the efficiency but not the accuracy of transcription initiation.

Translational efficiency is thought to be less important in eukaryotes than in prokaryotes in determining relative protein abundance as most of the control of protein levels is thought to occur at the prior steps of transcription, mRNA processing and transport of mRNA to the cytoplasm (Reznikoff and Gold, 1986). This has been found to be the case apart from altering the AUG codon or preceding it with another AUG.

While eukaryotes do not use a Shine/Dalgarno-type of interaction between the mRNA and the ribosome to select the initiation codon, there is some bias in the nucleotide frequencies used around initiation codons (Kozac 1984) . The largest bias is at position -3, where the residues are nearly always purines and 80% As. In prokaryotes the -3 position is also mostly As; (Gold *et al.*, 1981; Stormo *et al.*, 1982). Thus there is a conserved pattern in ribosomal recognition of initiation sites. Biases in nucleotide frequencies also exist in other positions and may contribute to the probability of using a particular initiation codon.

Besides being capped at the 5' end, eliminating excessive secondary structure is generally expected to increase the rate of translation of a particular protein. Fusing a coding region of interest to the leader sequence from a highly expressed gene can be used to ensure efficient binding at the 5' end. Several examples exisin which secondary structure 5' of or including, an AUG inhibits initiation (Kozak, 1984; Grens and Scheffler, 1990; Fu *et al.* 1991; Liebhaber *et al.*, 1992).

In the scanning model of initiation the primary determinant of translational efficiency at the first AUG of an mRNA is the binding of the ribosome to the 5' end of the message. The cap site includes the bases coding for the 5' terminal nucleotides of the mRNA in this region. The structure of the mRNA influences the dependence of on the cap of ribosome binding: less structured messages are less affected by elimination of the cap (Kozac, 1983).

Large deletions of the 5' untranslated region can sometimes have negligible effects on translation (Spindler and Berk, 1984), while some data has indicated the amount

of poly A on the 3' end of the mRNA influences the rate of initiation (Palatnik *et al.*, 1984). Deletion of the cap site element usually results in decreased RNA chain initiation at new sites 30 nucleotides approximately from the TATA box (reviewed in Corden *et al.*, 1980; Breathnach and Chambon 1981, Manley, 1983). *In vitro* studies have shown that at least one factor interacts specifically with the TATA box and cap site regions of the DNA to form a stable preinitiation complex (Davidson *et al.*, 1983). This interaction occurs in the absence of RNA polymerase II and is stabilised by a factor with actin-like properties. At this stage the promoter is committed, so that subsequently other factors and RNA polymerase can specifically initiate RNA synthesis (Matsui *et al.*, 1980; Fire *et al.*, 1984).

Certain bases flanking the AUG affect its efficiency as an initiation codon. Kozac (1984)identified the sequence ACCAUGG as the eukaryotic consensus sequence for initiation of translation from a group of 211 mRNA sequences from plants and animals. Lutke *et al.* (1987) furthermore conducted structural and functional analyses of the consensus sequences surrounding AUG initiation codons. They proposed that factors which select AUG initiation codons in plants and animals differ significantly and identified the most common bases flanking start codons in plant mRNAs, these being the most efficient start codon context sequences. The consensus sequence for plants was AACA**AUG**GC. The AACA plant leader was included in the KY sequence design for plant expression (Chapter 2).

As in various prokaryotic systems characterised, highly expressed genes in plants are more restricted in their codon usage than plant genes in general. Murray and coworkers (1989) undertook a study of 207 plant genes, 53 monocot and 154 dicot, and found the codon dialect for plant genes differ between these two groups primarily in the use of G and C in the degenerate third base. There is also an indication that the tRNA populations of specific plant tissues may be adapted for optimum translation of highly expressed proteins such as zein.

Perlak and coworkers (1991) found that modifying the codon usage of the *cryIA(b)* and *cryIA(c)* genes of *Bacillus thuringiensis* in plants, enhanced their expression over 100-fold. This was found after the use of different promoters, fusion proteins

signals and environmental stressors such as wounding and pathogen invasion, result in a dramatic upregulation of protein sorting in the secretory system. This occurs, for example, on fungal or pathogen attack, when different defense molecules, such as isozymes of chitinases or β-glucanase are synthesised simultaneously. The isozymes that contain vacuolar sorting determinants are directed to the vacuole, while others are secreted to the cell wall, secretion being the default pathway in plant cells, as in bacteria (Dorel *et al.* 1988).

The mechanism by which seed storage proteins are synthesised, targeted and sequestered for long-term storage in endosperm cells has attracted much interest. There is now convincing evidence apart from the role of a signal peptide for ER translocation, for either a propeptide or internal sorting domains in vacuolar targeting of these storage proteins (reviewed by Bednarek and Raikhel, 1992). In certain cells, there is elevated activity by the secretory system, in which protein traffic is more intense, as in the storage parenchyma cells of developing seeds, in which more than 50% of the newly-synthesised proteins enter the secretory system and accumulate in protein storage vacuoles. These include seed storage proteins and plant defence proteins such as lectins and enzyme inhibitors (Dorel *et al.*, 1989). Soluble vacuolar proteins and those destined for the tonoplast itself are targetted by different mechanisms (Gomez and Chrispeels, 1993). The molecular signalling mechanisms controlling sorting are also becoming better understood (Stack *et al.*, 1993).

It is generally accepted that the Golgi body plays a central role in the intracellular sorting and delivery of polypeptides carrying the appropriate targeting signal to the vacuole. The vacuole targeting signal may be at the N-terminus, the C-terminus, or within a region of the mature protein itself (Neuhaus *et al.*, 1991). A plant vacuolar processing enzyme has been characterised (Hara-Nishimura *et al.*, 1993). From the many studies done, it is evident that, as in yeast, a number of different, unrelated sequences may carry out vacuolar targeting in plants (Chrispeels and Rhaikel, 1992). In the present study, the prevalence of signal sequences involved in targeting of proteinase inhibitors was considered in the design of a KY expression construct for expression of the inhibitor in tobacco (4.1.2).

Potato and tomato plants contain two small multigene families that encode two serine PIs called inhibitors I (C) and II (T/C). Members of both gene families are expressed in leaves in response to chewing insects or other mechanical wounding. These are in many cases targeted to the vacuole. Tomato PI I has both pre- and pro-sequences (Cleveland *et al.*, 1987; Graham *et al.*, 1985) and tomato PI II and potato PI II have pre- sequences (Graham *et al.*, 1985). After post-translational modification, the mature proteins are sequestered in the vacuole (Walker-Simmons and Ryan, 1977b; Johnson *et al.*, 1989). At least three different PIs have been identified in soybean. These include the Kunitz trypsin inhibitor and BBPI (T/C) and a family of its isoinhibitors (Hwang *et al.*, 1978). The soybean Kunitz trypsin inhibitor of soybean (*Glycine max* L) has a 25 amino acid hydrophobic N-terminal signal sequence (cleaved at Asp/Phe). A genomic clone encoding the isoinhibitor C-II and a partial cDNA for BBTI has been characterised (Hammond *et al.* 1984) as well as cDNA clones for isoinhibitors D-II and C-II from soybean (Joudrier *et al.*, 1987). A full-length cDNA clone of the BBPI of soybean has been found to have a signal peptide of 39 amino acids (Beak and Kim, 1993). Maize seeds contain a 12 kD trypsin inhibitor, selective in its inhibition of activated Hageman factor, which is referred to as CHFI for corn (activated) Hageman factor inhibitor. A cDNA clone of CHFI revealed a stretch of 28 amino acids at the N terminus not found to be present in the inhibitor isolated from seeds, very similar to the sequences of other cereal endosperm proteins. The putative site of cleavage of the signal is Ala/Ser (Wen *et al.* 1992). Elpidina (1991) reports the localisation of a metalloproteinase and its inhibitor in the protein bodies of buckwheat seeds. These findings are further discussed in 4.1.2.

Plant PIs have complex expression patterns, being produced both constitutively (having developmental and tissue-specific expression) and in response to wounding and pathogen attack (Rosahi *et al.*, 1986; reviewed in Ryan, 1988; Stanford *et al.*, 1989; Keil *et al.*, 1989; Sanchez-Serrano *et al.*, 1993; Davis *et al.*, 1993; Hollick and Gordon, 1993). Regulation of a heterologous gene is an important consideration in cloning studies (Gallie, 1993) and this was investigated for the present study by referral to the workof Gatehouse *et al.*. (1986) and Ryan (1988). The

studies into the regulation of PIs expressed in transgenic plants have greatly aided elucidation of the wound-induction signal pathway in plants (Davis *et al.*, 1988; Xu *et al.*, 1993). In many cases wounding has been found to result in systemic induction of PIs (Parsons *et al.*, 1989, Palm *et al.*, 1990). Pearce *et al.*, (1991) found a polypeptide from tomato leaves induces wound-inducible PI proteins. This hormone peptide has been been named systemin and identified as the long-sought proteinase inhibitor-inducing factor, PIIF (Ryan *et al.*, 1992). Other wound induction signals found to be involved in induction of PI expression include abscissic acid (Pena-Cortez *et al.*, 1989), jasmonic acid (Yamagishi *et al.*, 1993), and auxin (Taylor *et al.*, 1993).

The identification of DNA control regions involved in wound induction and of nuclear factors binding to these regions have shed light on the developmental and environmental regulation of PIs (An *et al.*, 1989; Lorberth *et al.*, 1993; Meller *et al.*, 1993; Habu *et al.*, 1993; Yu *et al.*, 1993). However, work in this field is still at an early stage and expression patterns from transgene fusions with wound-inducible promoters are variable. No tissue-specific or wound-inducible promoter was thus chosen for KY expression. A second consideration mitigated against the use of a wound-inducible promoter for KY. This was the report of an interesting example of insect feeding behaviour amongst the Lepidoptera which seems to prevent the systemic induction of the wound response. This feeding behaviour was observed as the insect chewing off the leaf on which it has latest fed at the petiole, (leaf docking). This is proposed to prevent the wound signal from being systemically transmitted to the rest of the plant (Edwards and Wanjura, 1989).

Many advances have been made recently towards optimising expression of heterologous constructs in transgenic plants. These include the utilisation of motifs found within the 5′ untranslated leaders of certain plant viruses (Gallie and Walbot, 1992; Dowson-Day *et al.*, 1993) and the use of multiple *cis* regulatory elements for maximal expression from CaMV 35S promoter constructs (Fang *et al.*, 1989). Fusion with a leader of a highly expressed protein can also be an effective strategy, as accomplished by Wong and coworker (1992). These authors used plant (*Arabidopsis thaliana*) leaders and transit peptides to enhance expression of

proteins of *Bacillus thuringiensis*.. One or a combination of these strategies (Sleat and Wilson, 1992) could be potentially used to further work in transfering pest resistance via cloning of a PI such as KY. Were the KY construct itself found to be insufficient to confer field resistance to crop plants, it may be advantageous to combine its expression with another inhibitor to confer adequate field resistance to crop plants (Gatehouse *et al.*, 1993). In this respect, synergistic effects between two-way combinations of inhibitors have been found (Burgess *et al* ., 1994).

The development and application of plant PIs as biocontrol agents may be said to be of great advantage in reducing the cost and pollution problems associated with chemical control methods used to date. One of the major advantages of the use of PIs includes that development of insect resistance to the control agent itself, which has occurred in the application of *Bacillus thuringiensis* endotoxin (Van Rie, 1991) is unlikely in the case of a PI as the inhibitor targets the active site of a major metabolic enzyme. The effect of the PI is furthermore specific, being targeted to those insects feeding on the plant while at the same time being of sufficiently broad host range to be effective in combating a range of insect pests.

CHAPTER TWO

PRELIMINARY EVALUATION OF THE KY PEPTIDE AS A BIOCONTROL AGENT AND CLONING KY SEQUENCE CONSTRUCTS INTO pBLUESCRIPT

CONTENTS

CHAPTER TWO

PRELIMINARY EVALUATION OF THE KY PEPTIDE AS A BIOCONTROL AGENT AND CLONING KY PEPTIDE SEQUENCE CONSTRUCTS INTO pBLUESCRIPT

2.1 INTRODUCTION

The advancement of molecular biology, especially genetic engineering, has greatly increased the use of synthetic nucleic acid fragments, in particular genes for biologically active peptides and proteins (Denefle, 1987; Chen *et al.*, 1990; Ciccarelli *et al.*, 1990; Ivanov *et al.*, 1990; Pierce, 1994). Three well established approaches to oligonucleotide synthesis are the phosphodiester, phosphotriester and phosphite triester methods (Efimov *et al.*, 1986). The use of synthetic sequences has in some cases greater potential than cloned cDNAs or genomic sequences isolated from natural sources. For instance it can allow the production of double-stranded polynucleotides of predetermined sequence, including specific design features such as conveniently situated restriction sites and transcription and translation regulatory signals. In addition codons favourable for enhancing synthesis of corresponding proteins in heterologous cells can be used. The synthetic approach simplifies directed mutagenesis and holds promise for obtaining analogues of a peptide or protein with alterations at functionally important positions.

Oligonucleotide synthesis is performed in the 3' to 5' direction on a controlled pore glass (CPG) column support. According to the phosphoramidite method, the chain is extended by successive additions of di-phosphoramidites to a starting nucleoside attached to the CPG by the 3' OH group (the CPG- nucleoside). Coupling of each phosphoramidite is preceded by detritylation of the previous nucleoside, which serves to deprotect or unblock the 5' end of the extending chain for further additions (Fig. 2.1). The steps involved in the synthesis form a cycle as outlined below.

a. Detritylation of the CPG-nucleoside, using 2% tri- or di-chloroacetic acid (TCA or DCA) in dichloromethane, to free the 5'-OH.

b. Activation of the next phosphoramidite, which requires a tetrazole group and creates a reactive nucleoside derivative.

c. Addition of the activated nucleoside to the reactive CPG-nucleoside obtained in

step a.

d. Capping of any unreacted CPG-nucleosides which still have a 5'-OH group, by acetic anhydride and diethyl amino pyridine (DMAP) to ensure no further reaction (extension of failure sequences).

e. Oxidation of the phosphate to a pentavalent phosphate, using iodine, tetrahydrofuran (THF), 2,6-lutidine and water.

This completes the cycle and finally the CPG-oligonucleoside is deprotected (unless the trityl group is to be used for purification procedures) and cleaved from the support.

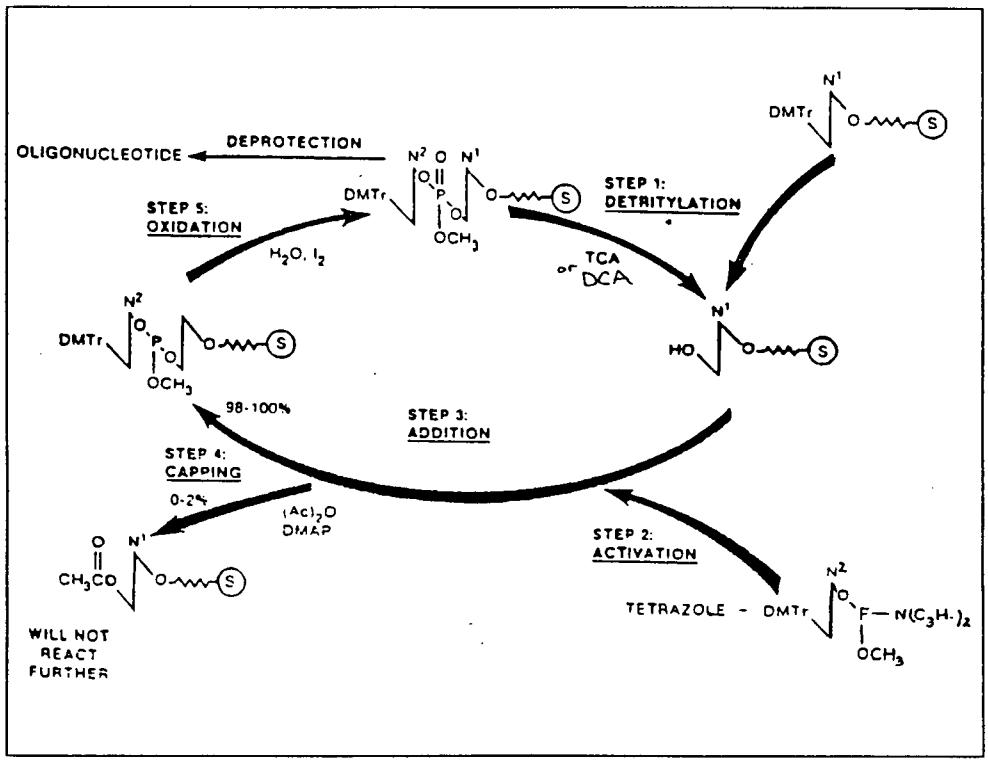


Fig. 2.1 Steps in the cycle of nucleotide synthesis.

The purification procedure for oligonucleotides synthesised in the Department of Biochemistry, University of Cape Town, uses the reverse-phase cartridge (RPC) method, with certain final steps altered to derive a rapid technique which uses n-butanol. The RPC method differentiates between the desired oligonucleotide and shorter truncated products using the hydrophobic properties of the 5'-dimethoxytrityl (DMTr) group, which is only present in the full-length oligonucleotide sequence. The 5' DMTr groups are subsequently removed from the oligonucleotide by treatment with 80% acetic acid and then precipitation of the product through ammonia and butanol.

The aims of this project were the design, synthesis and cloning of sequences coding for the improved analogue of the antitryptic loop of the DE3 inhibitor, KY, to enable expression in plants and in bacteria. Ultimately, bioassays would be needed to establish the efficiency of this novel inhibitor, and the range and degree of susceptibility of target insect pests.

This chapter describes the achievement of the initial aims. The potential value of using the KY peptide as an inhibitor for pest control was first assessed in preliminary feeding trials using the synthetic peptide. Once this was indicated, DNA sequences encoding KY were designed and synthesised for cloning into the pBluescript vector.

2.2 MATERIALS AND METHODS

All methods not referenced or described here were according to Sambrook *et al.*, 1989.

2.2.1 Bioassays to assess insecticidal activity of synthetic KY peptide.

Peptides, supplied by the Department of Biochemistry, University of Cape Town, were synthesised on an automated solid phase peptide synthesizer (Milligen 9050 Automatic Pepsynthesiser). Purification was firstly by high pressure liquid

chromatography (HPLC), using a Millipore Waters pump system with a C_{18} NOVA-PAK 8 mm X 10 cm cartridge. The elution solvent was 0.1% trifluoroacetic acid (TFA) in Milli-Q water with a 5-25% acetonitrile gradient, used at a flow rate of 1 ml/min (Fig. 2.2).

The peptide was eluted isocratically under the following gradient conditions: Buffer A: 0.1% trifluoroacetic acid (TFA) in Milli-Q water; Buffer B: 0.1% TFA in acetonitrile; 0 min 95% A, 5% B; 30 min 75% A, 25% B. Cyclisation of the peptide was done by oxidation according to the method described in Sunde (1989).

Preliminary bioassays were conducted at the South African Sugar Experiment Station (SASEX), Natal, utilizing artificially synthesized purified KY peptide. The HPLC-purified peptide product was incorporated into the rearing medium fed to second instar larvae of a major sugarcane pest, the stem-borer *Eldana saccharina* Walker in controlled feeding trials. Two tests were done in which inhibitor was incorporated into artificial diet to a final concentration of 0.01 %. Larvae were placed onto the diet and monitored weekly for abnormal development over four weeks.

2.2.2 Design and synthesis of four sequences encoding the trypsin inhibitor KY

Artificial constructs were designed for expression in both bacteria (*E. coli*) and plants. The amino acid sequence of the altered 11-amino acid peptide was reverse-translated using published bacterial (Konigsberg and Godson, 1983) and plant (Murray *et al.*, 1989) codon usage. The relevant transcription and translation initiation and termination consensus sequences were used to optimise expression in each.

DNA, RNA and protein sequence analysis packages used were from the Genetics Computer Group (GCG; University of Wisconsin) package version 6.1 run on a DEC/VAX 6000-330 mainframe computer. The DNA sequences were scanned for restriction endonuclease recognition sites with the MAP program and also for unplanned stop codons. As excessive secondary structure can halt or decrease

through a 10% Hydrolink gel matrix (HL-D600; AT Biochem; Appendix), to confirm the concentration and strandedness of the oligonucleotides before annealing.

Complementary strands were paired at a ratio of 1:1, and annealed under four different conditions, two of temperature and two of ionic concentration. Aliquots of 5 µg of each oligonucleotide were suspended in a solution of 100 µl final volume in which ionic conditions were established by addition of either 1 M NaCl to a final concentration of 100 mM, or 5X Rx buffer (100 mM $MgCl_2$, 250 mM NaCl, 200 mM Tris-HCl, pH 7.5; Sequenase^R kit, US Biochemicals). Samples were annealed at 45°C and 65°C for two minutes, in a water bath followed by slow cooling for 30 min as above.

The efficiency of the annealing reaction was assessed by a method used for spectrophotometric determination of the temperature of dissociation of DNA templates (T_m value). This was obtained from a plot of the melting curve of each template sample, using a thermostatted Beckman Diode Array Spectrophotometer with a heating block. Absorbance readings were taken from A_{260}-A_{280} every 20 seconds over 1 h as the temperature of the sample in the cuvette was increased from 30°C to 90°C.

Annealed forms of the oligonucleotides were additionally differentiated by polyacrylamide gel electrophoresis (PAGE). Strandedness of annealed products was assessed on vertical 10% gels, (227mm X 117 mm X 1 mm) using a gel mix for low molecular weight DNA, (Hydrolink HL-D600; AT Biochem), run at 200 V for 4 hours. 500 - 1 000 ng DNA were loaded per well in 5-10 µl Orange G tracking dye (0.06% Orange G, 5% sucrose).

The pBluescript vector was used (Stratagene, La Jolla, CA), having the advantages of multiple cloning sites and ease of sequencing (Fig. 2.2). Linearised vector, pSK was purified by electrophoresis through 0.7% Tris-acetate EDTA (TAE)-buffered preparative gels and recovered from gel slices by the Gene-clean™ method according to the manufacturers' instructions (Bio101, USA). In another method,

the 5' termini of vector DNA were dephosphorylated by treatment with calf intestinal phosphatase, (CIP) in order to prevent self-ligation of vector termini. The DNA was resuspended in a volume of 35 µl of sterile distilled water, 4 µl of (10X) CIP buffer (10 mM $ZnCl_2$, 10 mM $MgCl_2$, 100 mM Tris-Cl (pH 8.3), and 5 IU CIP enzyme, and incubated at 37°C for 30 min. CIP was heat-inactivated at 75°C for 15-20 min, and the DNA phenol-cleaned and re-precipitated. The DNA was resuspended in distilled water to yield a concentration of 250 ng/µl.

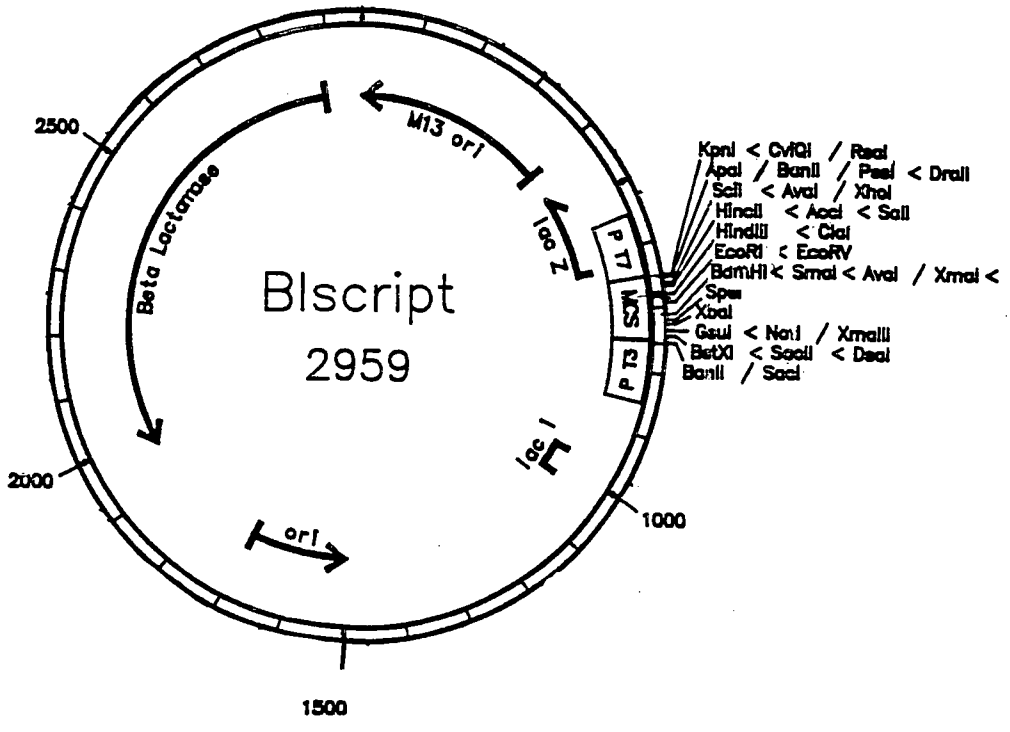


Fig. 2.2 Schematic representation of the Bluescript pSK+ vector, showing restriction sites used for cloning (*Cla*I and *Bam*HI) and screening (*Kpn*I and *Sac*I) KY constructs. MCS: Multiple cloning site, P T7 and PT3, promoters T7 and T3, ori, origin of replication, lacZ, beta-galactosidase gene, M13 origin of replication

In cases in which the vector was treated with calf intestinal phosphatase (CIP), the oligonucleotides were treated with polynucleotide kinase (PNK) to enable ligation to the vector, as synthetic oligonucleotides lack a 5' phosphate group. For the

addition of the 5' phosphate group to oligonucleotides, 5 IU T4 polynucleotide kinase (PNK), was added to 5 µg of oligonucleotide together with 1 µl ATP (at 100 mM) and 3 µl PNK buffer (50 mM Tris-Cl, pH 7.6, 100 mM $MgCl_2$, 50 mM dithiothreitol (DTT), 0.5 mg/ml bovine serum albumin, 1 mM spermidine and 1 mM EDTA), in a final reaction volume of 30 µl. The reaction was incubated at 37°C for 1 hour and stopped by addition of 1 µl 0.5 M EDTA.

Ligation volumes were initially kept dilute (4.6 ng/µl of pSK in 50 µl) to reduce the likelihood of concatemerisation of the annealled oligonucleotide product. Insert: vector molar ratios used w.r.t. picomole ends were 1:1 (4 ng insert: 238 ng SK DNA in a 20 µl ligation volume) or 2:1 and subsequently 5:1 and 10:1. The ligation mix was used to transform cells of *E. coli* LKIII made competent by the calcium chloride method and plated on LA plates containing 100 mg/ml ampicillin, 40 mg/ml X-gal and 5 mM IPTG. Other host strains used for transformation were *E. coli* LK112 (*recA*$^-$), JM109 and JM105.

2.2.4 Selection of recombinant clones

Colonies were first selected by their ability to grow on Ap^{100} plates and to α-complement lacΔM15 host cells on X-gal, IPTG-supplemented plates, to allow for blue/white colour selection. From these, *Dde*I restriction analysis of "miniprep" DNA was used for screening. Dot-blot hybridisation replaced this method.

Mini-preparations of DNA were screened by probing dot-blots on Hybond N$^+$ nylon filters according to a method adapted from the suppliers' manual (Amersham). Labelling of probes was performed at 37°C by addition to 500 ng of oligonucleotide of 20 IU T4 polynucleotide kinase (PNK), 3 µl 10X PNK buffer (100 mM $MgCl_2$, 50 mM dithiothreitol (DTT), 0.5 mg/ml BSA, 1 mM spermidine, 0.5 M Tris-Cl (pH 7.6), 1 mM EDTA), and 50 µCi [γ^{32}P] dATP (specific activity

~3000 Ci/mol) in a total reaction volume of 30 µl. The labelling mix was held at 37°C for 1 h and the reaction stopped by the addition of 2 µl of 0.5 M EDTA, pH 8.0.

In initial experiments, unincorporated [α-^{32}P]dATP was separated from labelled DNA by Pasteur column chromatography on Sephadex G-50. In later experiments, this step was found to be unnecessary, blocking and post-hybridisation washes being sufficient to remove background, non-specific hybridisation. Plasmid DNA was isolated from positive colonies, the fragment across the MCS of each excised with the enzymes *Kpn*I and *Sac*I and digests electrophoresed through vertical 7.5 % Hydrolink gels for 3 hours at 200V. The nucleotide sequence of recombinant plasmids with the appropriate sized fragments was determined according to a modification of the dideoxy chain termination method of Sanger *et al.* (1977), using the Sequenase version 2.0 kit obtained from U.S. Biochemicals (Tabor and Richardson, 1987).

2.3. RESULTS AND DISCUSSION

2.3.1 Bioassays

The synthetic KY peptide product obtained after HPLC purification (Fig. 2.3) was used in preliminary feeding trials. In the first feeding trial a 70% mortality rate was obtained while the control mortality was 6%. In the second, 28% mortality occurred and control mortality was 0%. Further tests were not possible due to the inavailability of more synthetic KY. However KY does appear to have an effect on the survival of second instar larvae of *Eldana saccharina* Walker. For an accurate quantitative determination of the LD_{50}, these trials would need to be repeated. Soybean trypsin inhibitor would be a good control for comparison.

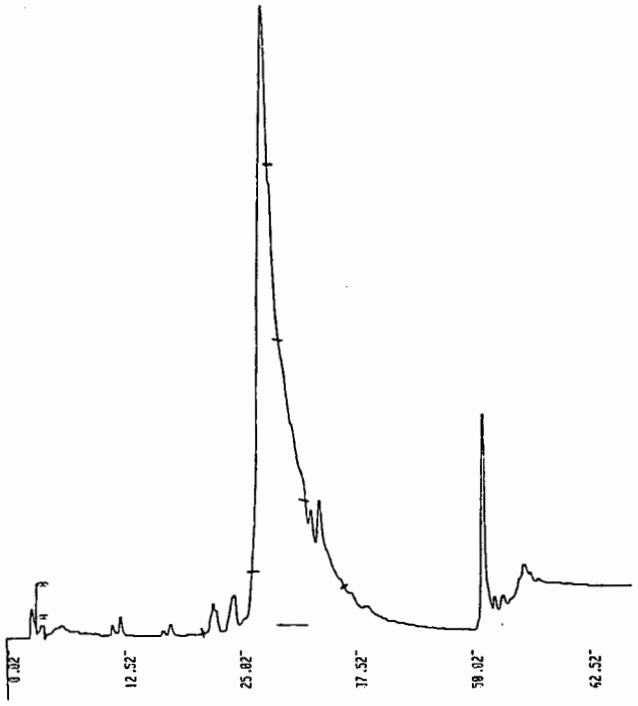


Fig. 2.3. HPLC elution profile of the KY peptide. The fraction containing the KY peptide is indicated by a small bar below the A_{229} peak.

2.3.2 Design, synthesis and preparation of oligonucleotides

For the bacterial sequence, a Shine/Dalgarno sequence and spacer (AGGAGG) were used (Fig. 2.4) and for the plant sequence, the sequence AACA, identified as a consensus sequence enhancing initiation of translation, 5' to the AUG of plant genes, (Lutke *et al.*, 1987; see 1.3.3) was included . Restriction enzyme sites *Cla*I and *Bam*HI were included at the 5' and 3' end of each sequence to allow for directional cloning. These sequences were named Bact-KY and Plant-KY respectively.

BACTERIAL-KY PEPTIDE SEQUENCE

(ClaI) Sp SD Met Ser Cys Thr Lys Ser Ile Pro Pro Gln Cys Tyr stop *BamHI*
CGAT AGGAGG AACACAA ATG TCT TGC AAC AAA TCT ATC CCG CCG CAG TGC TAC TAA GG
TA TCCTCC TTGTGTT TAC AGA ACG TTG TTT AGA TAG GGC GGC GTC ACG ATG ATT CCT AG

BACTERIAL-KY-TUFTSIN FUSION PEPTIDE SEQUENCE, COOH TERMINUS

Thr Lys Pro Arg stop *(BamHI)*
ACC AAA CCG GCT TAA G
TGG TTT GGC CGA ATT CCTAG

PLANT-KY PEPTIDE SEQUENCE

(ClaI) Sp Met Ser Cys Thr Lys Ser Ile Pro Pro Gln Cys Tyr stop *BamHI*
CGAT AACA ATG TCT TGC ACC AAG TCT ATT CCA CCA CAA TGC TAC TAA G
TA TTGT TAC AGA ACG TGG TTC AGA TAA GGT GGT GTT ACG ATG ATT CCT AG

PLANT-KY-TUFTSIN SEQUENCE, COOH TERMINUS

Thr Lys Pro Arg stop *(BamHI)*
ACC AAG GCA CGT TAA G
TGG TTC GGT GCA ATT CCT AG

Sp: spacer
S/D: Shine-Dalgarno

Fig. 2.4 Nucleotide and amino acid sequences of the four KY constructs designed for cloning into pBluescript, pSK$^+$.

As the small size of the peptide would limit immunogenicity and as it has been shown that protease inhibitors are poorly immunogenic, two further constructs were designed to encode the highly immunogenic tetra-peptide, tuftsin at the 3' end of the sequence (Dagan *et al.*, 1987). Tuftsin (Thr-Lys-Pro-Arg) is an immunoglobin heavy chain-associated tetrapeptide which augments the antigen-presenting capacity of macrophages *in vitro* when supplied simultaneously with the antigen. The coding sequence for tuftsin in each construct was placed after the Tyr codon, the last amino acid of KY and immediately prior to the stop codon (Fig. 2.4). These sequences encode fusion proteins of 15 amino acids for expression in plants and bacteria respectively. The product KY-Tuftsin would be immunologically assayable *in vitro* and thus potentially useful for determining

the production and correct folding of translation products in recombinant bacteria and/or transgenic plants. Antibody raised against the peptide could be used to establish this by immunological analysis of cell extracts.

The ΔG values of the mRNAs of the Bact-KY and Plant-KY sequences indicated they would not form very stable structures, and thus translation would not be inhibited (Fig. 2.5).

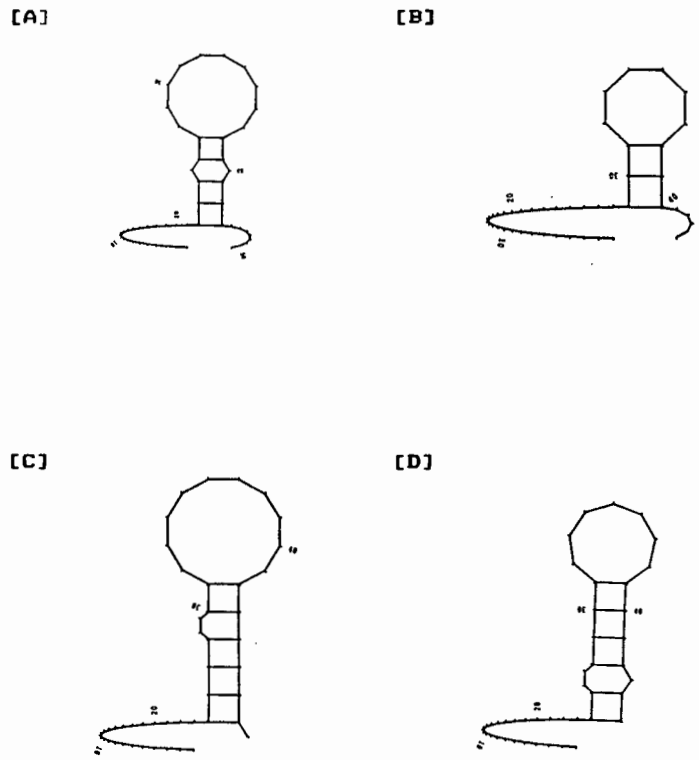


Fig. 2.5 Theoretical folding structures of the mRNA of KY sequences and their stability as predicted by Fold (GCG). [A] Bact-KY mRNA, ΔG = -2.4 kCal/mol, [B] Plant-KY mRNA, ΔG = -0.8 kCal/mol, [C] Bact-Tuft mRNA, ΔG = -3.8l kCal/mol, [D] Plant-Tuft mRNA, ΔG = -3.4 kCal/mol.

The pH of the oligonucleotides which had not been cartridge cleaned was checked and they were found to be at pH 5. Around this pH, DNA precipitation and the risk of oligonucleotide degradation begins. These oligonucleotides had evidently

precipitated out of solution as determined from O.D. readings at 260 nm taken before and after butanol precipitation. This explained the failure to obtain recombinant clones in pSK+ or to obtain dissociation curves expected from temperature shift experiments used to assess the efficiency of annealing. This problem was solved by butanol/ammonia treatment of the oligonucleotides. This step may also have eliminated any oligonucleotide fragments which could have degraded at the low pH. Butanol precipitation furthermore rids the DNA of benzamite residues which can be present in the preparations post-synthesis. After butanol precipitation, single stranded oligonucleotide preparations were confirmed to be of the correct concentration on 10% polyacrylamide gels (results not shown) after which work with annealing continued.

2.3.3 Annealing reactions

The presence of annealled oligonucleotide products was firstly identified spectrophotometrically, from a plot of the sample undergoing the hyperchromic shift, which indicated melting of duplex DNA (Fig. 2.6). This provided clear evidence of the presence of annealed DNA. Annealing was found to have been successful in both buffers used.

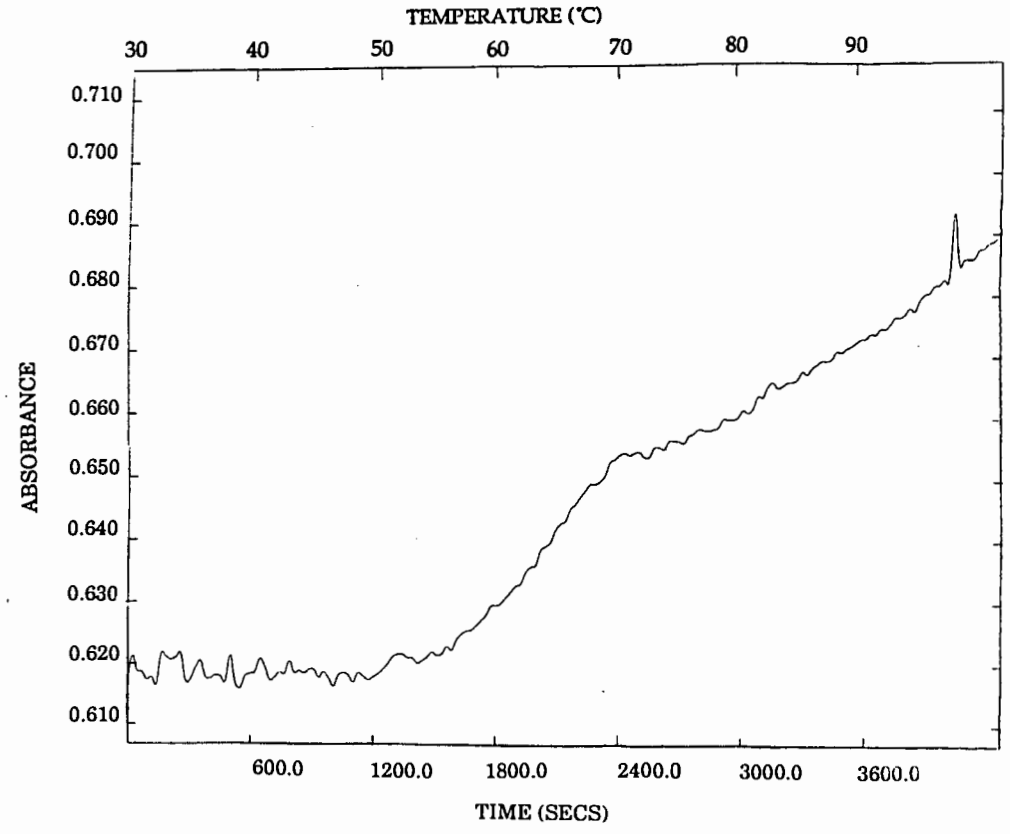


Fig. 2.6 Spectrophotometric detection of the dissociation of annealled oligonucleotide strands in Rx buffer upon melting. At time zero, the temperature was 30°C, increasing incrementally to 90°C by 3600 sec. Changes in absorbance readings, from A_{260}-A_{280} are indicated as the temperature of the sample in the cuvette increased over 1 h. The hyperchromic shift obtained from the annealled Plant-KY template indicates the dissociation of the two strands and thus the presence of duplex DNA.

Annealled oligonucleotide products were observed to migrate more slowly than their single-stranded counterparts on 10% Hydrolink gels. This method yielded both qualitative and quantitative evidence of annealing (Fig. 2.7). While there were two bands seen from annealed Bact-KY and Plant-KY templates, it was reasoned that only correctly-annealed products could participate in the cloning event and it was decided to proceed with ligations and optimised screening steps.

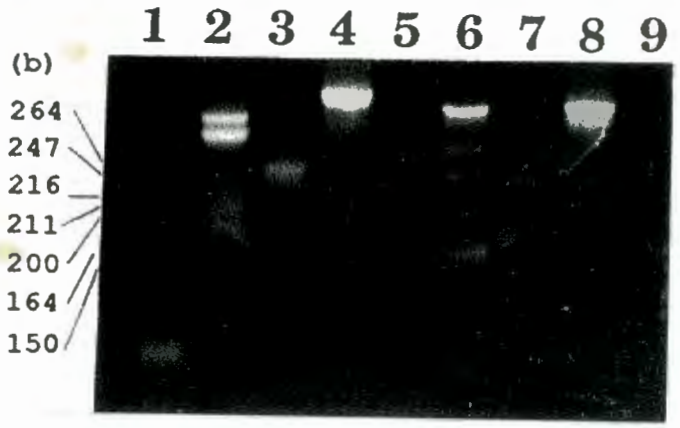


Fig. 2.7 PAGE analysis of annealled oligonucleotide products on 10% Hydrolink gel indicating size differences between ss and ds forms before and after annealing. Lanes 1 and 2, ss and ds Bact-KY; 3 and 4, ss and ds Bact-KY- Tuft; 5 and 6, ss and ds Plant-KY; 7 and 8, ss and ds Plant-KY-Tuft. The ds products form fragments of 54; 59; 48 and 63 bp respectively.

2.3.4 Cloning the KY sequences into *E. coli*

A number of variables were changed and methods added during the cloning procedure to optimise conditions in order to obtain the clones of interest (Fig. 2.8A). In optimising the cloning strategy additional levels of screening were devised to detect the presence of both annealed products and recombinant clones.

The KY sequence causes a frameshift in the N-terminus of the *lac*Z reading frame and would therefore be expected to enable identification of the recombinants with this sequence by blue/white colour selection even if there were readthrough through the KY sequence. However, false white colonies were obtained and colonies were obtained which were both blue and white upon streaking onto LA X-gal plates.

Screening by restriction analysis of miniprep plasmid DNA from white colonies with the enzyme *Dde*1 was also misleading. The fragment sizes obtained from partial digests of vector alone (1115 and 706 bp) yielded restriction patterns similar

to those expected if *Dde*I had cut at the single site in the KY sequence in the recombinant plasmid (1185 bp and 745 bp). Although manipulations such as further miniprep clean-up (phenol treatment etc.) were possible, this method was deemed unreliable for screening colonies.

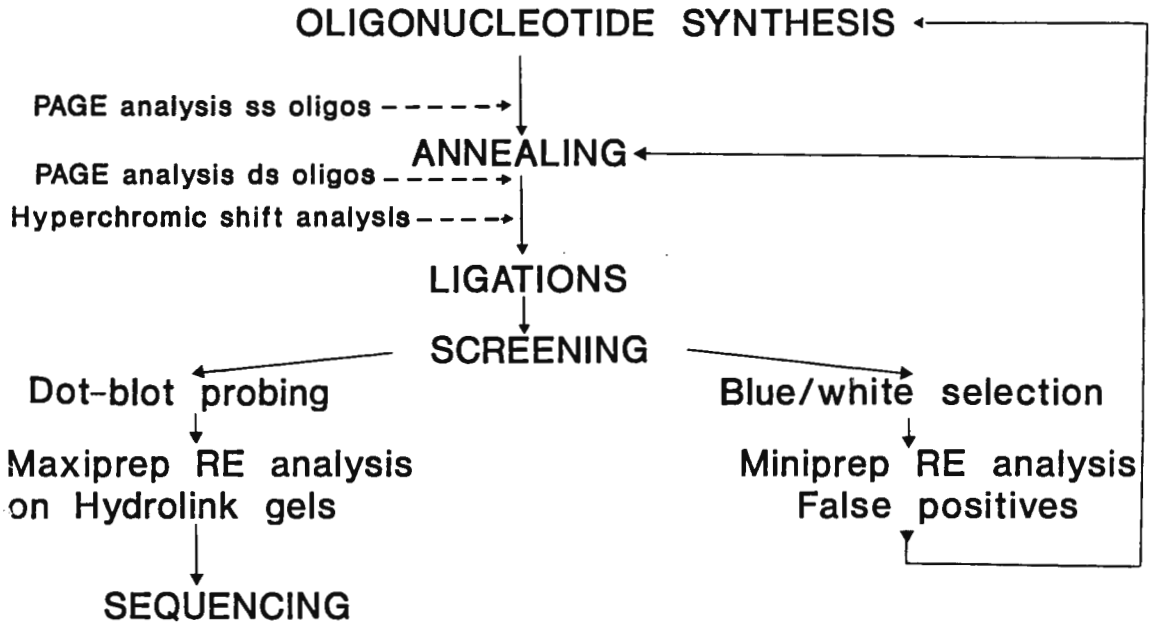


Fig. 2.8 Parameters and screening methods optimised in cloning the KY sequences into pBluescript SK+

Subsequently, dot-blot hybridisation was used for screening. However, no positives being obtained, the success of the annealing step was analysed. As no evidence of annealing was found at this stage, the oligonucleotide preparations were checked and were found to require butanol treatment. Checks introduced to follow the results of the annealing reactions more closely were helpful in reducing uncertainty w.r.t. strandedness of the preparations and time-saving in that ligations could proceed with DNA of the correct concentration. Gene-cleaning was found to be effective in the elimination of parental DNA, including singly-cut molecules and uncut vector molecules. De-phosphorylation of the vector by CIP treatment before ligation was a further method used to eliminate parental DNA molecules arising from religation of vector cut at one site only. Subsequent changes included kinase treatment of the oligonucleotides and

decreasing the volume of the ligations to 20 µl. The latter was chosen as it was considered insert-insert ligations would be rare.

However, the above factors were not likely to have been as critical in obtaining the clones of interest as the butanol precipitation step, as after this step clones were obtained directly, using a simplified approach without CIP treatment of the vector or kinase treatment of the oligonucleotides. Both white and pale blue colonies were selected from these transformations, the latter observed to have blue/white streaking patterns when grown on X-gal, IPTG replica plates. Random colonies with both these phenotypes were screened by dot-blot hybridisation and a high percentage found to be positive (Fig. 2.9).

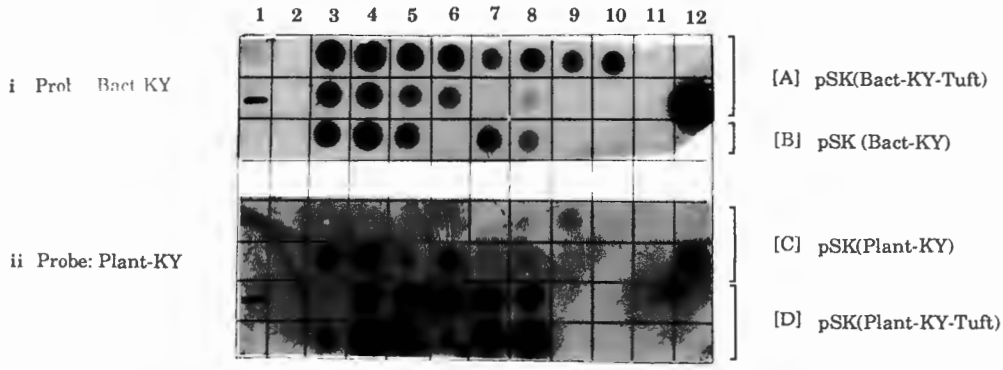


Fig. 2.9 Dot-blot hybridisation identifying recombinant plasmids containing KY-derived sequences from miniprepared DNA using [γ-^{32}P]dATP end-labelled oligonucleotides as probes. (i) Positive control (row 12), 50 ng Bact-KY oligonucleotide (5'-3'). (ii) Positive control (row 12), 50 ng Plant-KY(5'-3'). Negative controls (-), pSK miniprepared DNA.

Plasmid DNA isolated from colonies identified from dot blots as presumptive clones were analysed by restriction analysis and PAGE. Excising the MCS polylinker of pSK as a *Kpn*I-*Sac*I fragment effectively enabled confirmation of the

cloning event on 10% Hydrolink gels. The sequences Bact-KY (54 bp); Bact-KY-Tuft (69 bp) and Plant-KY-Tuft (63 bp), including the 68 bases contributed by the *Kpn*I-*Sac*I fragment of pSK would produce 122-mer, 137-mer, and 130-mer fragments respectively. The expected pattern from these fragments was evident (Fig. 2.10A).

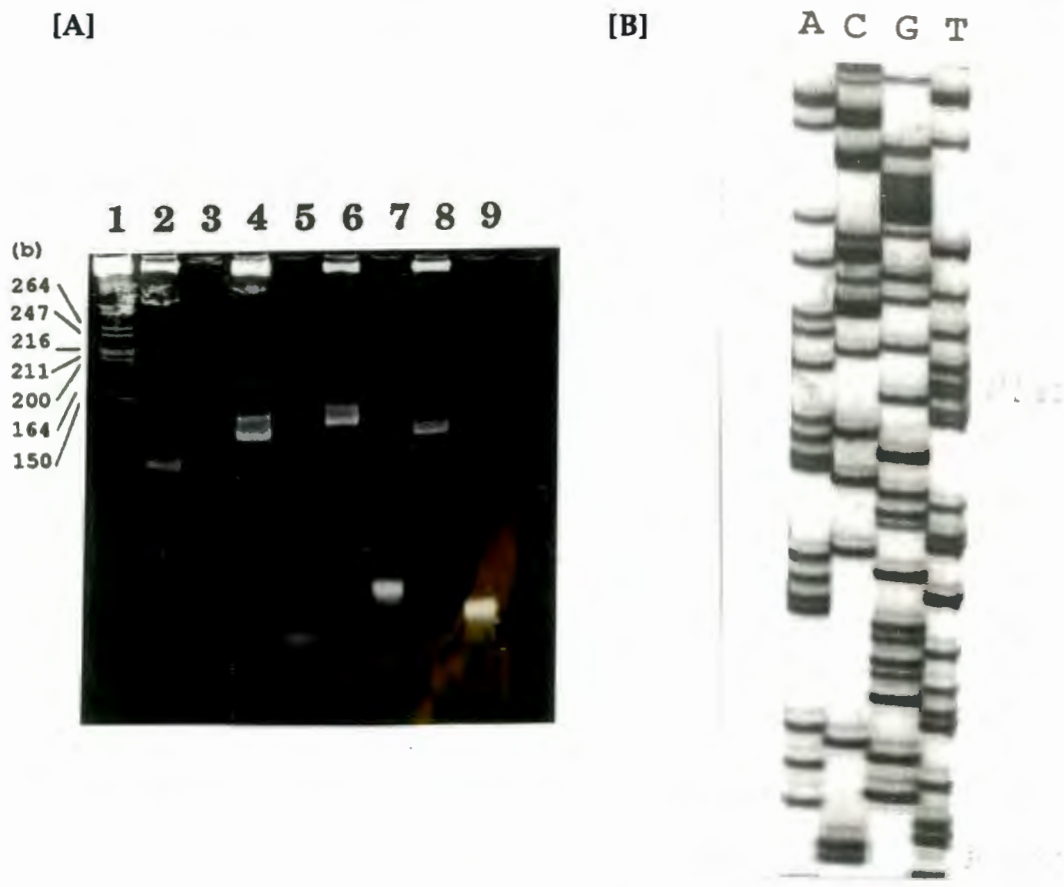


Fig. 2.10 **[A]** Confirmation of KY clones by restriction analysis and PAGE on 7.5% Hydrolink. Lane 1, λ(*Pst*I); 2, pSK(*Kpn*I/*Sac*I); 3, pSK(Bact-KY) (*Kpn*I/*Sac*I); 4, Bact-KY oligonucleotide; 5, pSK(Bact-KY)(*Kpn*I/*Sac*I); 6, Bact-KY oligonucleotide; 7, pSK(Plant-KY-Tuft)(*Kpn*I/*Sac*I); 8, Plant-KY oligonucleotide.
[B] Sequence analysis of the Plant-KY construct

The Plant-KY fragment was identified similarly (results not shown). Recombinant plasmids found to contain fragments of the sizes of interest were sequenced in order to confirm that no mutations had occurred during or after synthesis. The acidity of the first preparations of oligonucleotides, which caused a

problem in their cloning can also cause depurination due to deprotonation of A and T residues, which can lead to mutations in the sequence. Fortunately, this concern was allayed after sequence confirmation of the clones as the four sequences were found to be correct (Fig. 2.10B).

CHAPTER THREE

CLONING THE KY SEQUENCE IN *E. coli* pET-12a

CONTENTS

CHAPTER THREE

CLONING THE KY SEQUENCE IN *E. coli* pET-12a

3.1 INTRODUCTION

The aim of the work towards the expression of the KY gene in *E. coli* was to enable a supply of recombinant peptide for feeding trials, further analysis of the peptide and possibly for the production of antibodies to KY free of plant antibodies.

The pET vector system developed by Tabor and Richardson (1985) and Studier and Moffatt (1986) is designed for optimal expression of recombinant proteins in *E. coli*. Target genes cloned in these plasmids are expressed in host cells which are lysogens of bacteriophage DE3, a lambda derivative carrying a DNA fragment containing the *lac*I gene, the *lac*UV5 promoter, the beginning of the *lac*Z gene, and the gene for T7 RNA polymerase, *rho*(Φ)10 (from gene 10 of T7 phage) inserted into the *int* gene. The T7 RNA polymerase is produced from its own translation start, and the gene is under control of the IPTG-inducible *lac*UV5 promoter. Addition of IPTG to a culture of a BL21(DE3) or HMS174(DE3) lysogen induces the polymerase, which then transcribes the target gene cloned downstream of the T7 promoter on the pET vector. T7 RNA polymerase is so active and selective that the target protein can comprise over 50% of total cell protein after a few hours' induction. The IPTG-inducibility of the T7 RNA polymerase, designed to be transcriptionally silent in the uninduced state and its target specificity, prevents loss of the plasmid if the cloned gene product is toxic to the cell.

The pET-12 vectors are designed to enable the secretion into the periplasm of recombinant protein by expression as part of a larger fusion protein with a signal peptide provided from the OmpT protease of *E. coli* (Fig 3.1). The 12 series provides for correct folding of the protein and *in vivo* removal of the leader at the natural cleavage site Ala-Ser. The pET-12a, b and c plasmids enable in-frame

cloning from three alternative sequential reading frames at the *Bam*HI site. The pET-12a vector was chosen to produce KY in order to promote export of the peptide to the periplasm to enable disulphide bridge formation (1.3.2). The pET-12 vectors include the T(Φ) transcription terminator just beyond the *Bam*HI site.

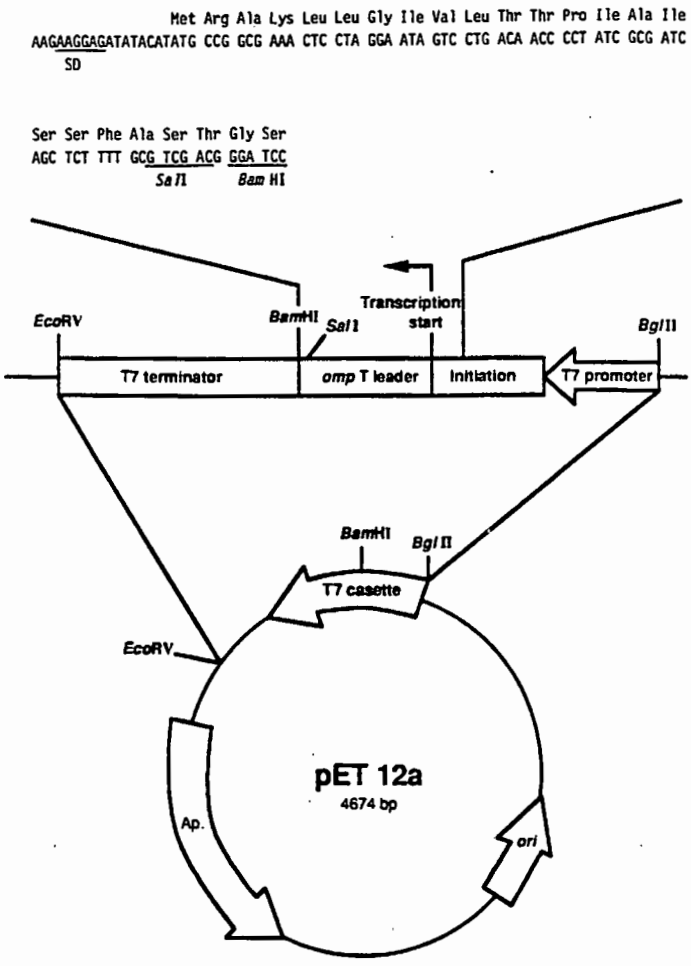


Fig 3.1 Schematic representation of the pET-12a expression vector. This contains the unique restriction sites *Bam*HI and *Sal*I between the T7 (Φ)10 promoter and the downstream T(Φ) terminator. Ap: ampicillin resistance marker.

Single-strand cloning by bridge-break repair

An adapted form of a method originally used for cloning single-stranded synthetic oligonucleotides for site-directed mutagenesis can be used for artificial gene cloning in *E. coli* (Mandecki 1986; Mandecki and Bolling, 1988). This

approach utilises the phenomenon of oligonucleotide-directed double-strand break repair in the cell. The procedure involves the transformation of *E. coli* with a denatured mixture of an insert-carrying oligonucleotide and linearised plasmid DNA (Fig. 3.2A). The transformed host bacterium supplies the second strand of the sequence annealed as a single strand into the blunted ends of a cut plasmid vector and repairs the break.

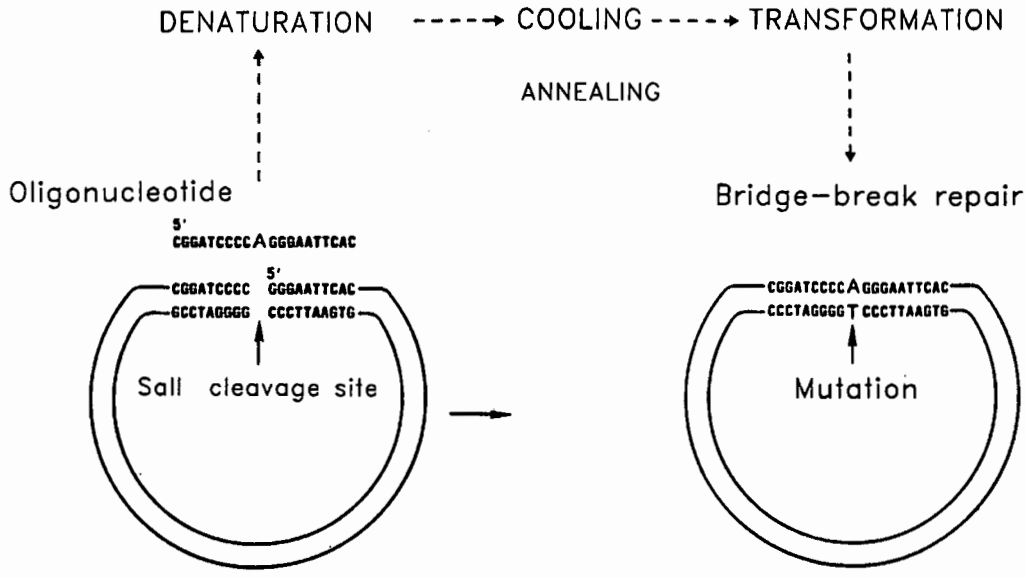


Fig. 3.2A Schematic representation of single-strand oligonucleotide-directed double-strand break repair cloning. The 'mutation' sequence represented by an A, carried by the oligonucleotide, can be accurately filled in up to 60 nucleotides (Mandecki, 1986).

The oligonucleotide contains the insert sequence bracketed by two 15-nucleotide arms which are complementary to the sequence of the cleavage site used to linearise the plasmid, and is presumed to assist in the formation of a gapped DNA hybrid structure. The repair of the gap occurs *in vivo* after transformation. A large molar excess of oligonucleotide over plasmid is required to enable effective cloning by bridge-break repair, viz. 1 000-fold molar excess of oligonucleotide (Mandecki, 1986).

In bridge-break repair cloning no enzymatic manipulation *in vitro* is required to effect ligation. The method is more than or at least equivalent to directional cloning as the homologous regions of the complementary arms extend further than a restriction site. Other advantages of this system include the low error rate (one in 4 000 bases cloned versus one in 160 for double-strand cloning (Mandecki, 1986; McClain *et al.*, 1986), and economy and ease of handling, as synthetic DNA for only one strand needs to be provided. The strategy thus provides an accurate, fast and simple method for the synthesis of a gene of a defined sequence.

3.2 MATERIALS AND METHODS

3.2.1 Cloning strategy for the expression of KY in *E. coli*

Mandecki (1986), in a range of annealing and transformation experiments with oligonucleotides of varying lengths established that to ensure adequate efficiency of cloning, oligonucleotides of a minimum length of 20-30 bases, preferably longer should be used. A high efficiency was observed (over 90%) using oligonucleotides 30 or more residues long. On the other hand, the accuracy of the fill-in reaction decreases as the distance between the left and right borders increases over approximately 60 nucleotides (Mandecki, 1986).

The construct was designed to enable cloning of the KY sequence as an in-frame fusion with the OmpT (outer membrane protein T) signal sequence of the pET-12a vector. A subclone of pSK-Bact-KY (2.3.4) could not be used to place the Bact-KY behind the leader sequence of pET-12a, due to the presence of the spacer and S/D sequence at the 5' end of this sequence. (The pET-12a expression vector became available only after constructs of chapter 2 were made.) A single oligonucleotide of 66 residues coding for the Bact-KY gene was designed and synthesized with left and right 15-bp arms having complementarity to either side of the blunted *Sal*I site of the pET-12a expression vector (Fig. 3.2B). Cloning into the *Sal*I site brings the codon of the first residue of KY, Ser into position adjacent to the last codon of the OmpT leader sequence, Ala (Fig. 3.2B).

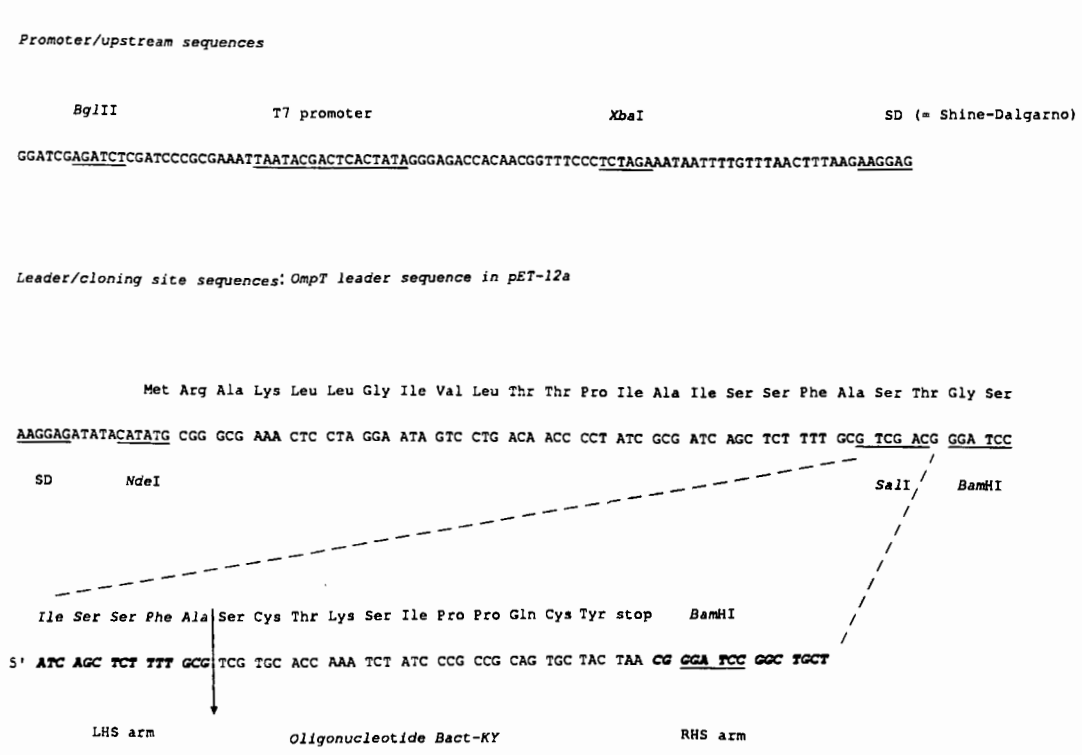


Fig 3.2B Design of the KY expression construct pET-12a-Bact-KY for oligonucleotide-directed double-strand break repair cloning. The vector, pET-12a, and oligomer sequences are shown. The 15-bp vector-homologous ends of the oligomer are in bold and italics. The oligomer Bact-KY anneals into the vector at the *Sal*I cleavage site and becomes the new sequence of the coding strand, the other strand being filled in during *in vivo* repair in the host cell. The arrow indicates the signal sequence splice junction of the OmpT leader, Ala/Ser.

Fortuitously this completes the natural cleavage site for the leader. This would allow for efficient *in vivo* cleavage of the signal by host cell signal peptidases, which recognise Ala-Ser as one of the the natural cleavage sites of the leader sequence from periplasmic proteins (Perlman and Havarlson, 1983). Only a slight alteration of one codon resulted, that encoding Ser, as the second amino acid in the splice junction in the pET-12a sequence is TCG instead of TCT of the Bact-KY sequence, the first amino acid of Bact-KY.

As unusual or excessive hydropathy can cause problems with cleavage of the leader peptide from the mature protein (Little *et al.*, 1989), the hydrophilicity of the recombinant fusion peptide from the construct was examined using the protein analysis programs Pepplot and PeptideSort (GCG, University of Wisconsin).

The Isoelectric program of the SAGA!! protein sequence analysis package (D.L. Maeder, University of Cape Town), calculating fractional charge at pH 7 with a window of one residue was used to show the charge distribution of the peptide. Amino acids with extreme pK' values (i.e. any abnormal distribution of charged residues) around the junction of the signal peptide up to 10 residues from the N terminus of the mature protein, is one of the major factors determining the efficiency of cleavage by signal peptidases (Summers, 1989b).

The causes of poor translation of some mRNAs can be factors such as an unfavourable distribution of rare codons or interfering structures in the mRNA (Studier *et al.*, 1990). These two characteristics were avoided in the design of the construct, in that high frequency codons were used (Konigsberg and Godson, 1983) and mRNA secondary structure was assessed using the Fold program of GCG (Zuker, 1984), which finds a secondary structure of minimum free energy based on published values of stacking and loop destabilising residues. This structure was plotted by Squiggles using an output file from Fold.

3.2.2 Single-strand cloning into pET-12a(OmpT) by bridge-break repair

The cloning strategy initially followed was linearisation of vector with *Sal*I, blunting with MBN, annealing of the single-strand oligonucleotide, and transformation of cells with the annealled mix. It was later found unnecessary to blunt with MBN. Enzymes and buffers used were purchased from Boehringer Mannheim.

Mandecki *et al.*, (1987) established optimal ratios of oligonucleotide and vector for annealing reactions. In the experiments of the present study, the amount of

oligonucleotide was similarly kept constant at 50 ng (20 fmols). The calculated correct molar amount of vector was 62 ng (20 fmols) and due to the inexact method of estimating vector concentrations after gel purification, three vector quantities were used in the denaturation and annealing reactions, viz. 30, 60, and 90 ng.

Vector and oligonucleotide were heated together for 3 min at 100°C, in 30 µl of denaturation buffer (10 mM Tris-HCl pH 8.0, 10 mM KCl, 5 mM $MgCl_2$), cooled at room temperature for 5 min and transformed into HMS174(DE3). *E. coli* HB101, a general purpose host for plasmid maintenance was used for maintenance and preparation of pET-12a plasmid.

Host strains for expression of target genes cloned in pET plasmids are lysogens of phage λDE3, which carries the T7 RNA polymerase gene under control of the IPTG-inducible lacUV5 promoter (Studier and Moffat, 1986). The *E. coli* pET expression host strain HMS174(DE3) (F^- *rec*A r^{k12-} m_{k12}^+ Rif^R (DE3) has higher competency than strain BL21(DE3) (F^- OmpT $r^{ß-}$ $m^{ß-}$) and was chosen for transformation experiments. BL21 is $RecA^+$ and has a somewhat lower transformation frequency. Furthermore, a few target genes are more stable in HMS174(DE3) or its derivatives than in the equivalent derivative of BL21(DE3) (pET manual, Novagen). This is thought to be due to amounts of some gene products inducing the SOS response in *E. coli*, leading to prophage induction, which kills the cells. The RecA deficiency of HMS174 prevents prophage induction and cell death.

The *E. coli* strain BL21(DE3)pLysE, which has tighter suppression of basal expression levels (Studier *et al.* 1992), was used as an expression host, and was transformed with the pET-12a-Bact-KY-6 construct. The single-stranded oligonucleotide encoding Bact-KY was re-cloned for this purpose, and the resulting plasmid transformed firstly into *E. coli* JM105 and then into BL21(DE3)pLysE.

Strains with pLysE carry the lysozyme gene under the control of the Φ3.8 promoter for T7 RNA polymerase. Lysozyme has dual activity in cutting a bond in the peptidoglycan of the cell wall of *E. coli* and in binding and inhibiting T7 RNA polymerase (Studier, 1991). Due to this inhibition of the T7 RNA polymerase, the presence of pLysE increases the tolerance of BL21(DE3) for toxic target plasmids. Plasmids that otherwise could not be established can be often be stably maintained and expressed in this strain. BL21 is also deficient in the *lon* and OmpT proteases.

Transformants were screened by dot-blot hybridisation with 5' end-radiolabelled probe, as in 2.2.4, or by restriction analysis of plasmids on 7.5% acrylamide gels. Changes used in attempts to obtain stable recombinant pET-12a-Bact-KY constructs were as tabulated below (Table 3).

Table 3. Parameters changed to obtain DNA from pET-12a clones.

Strains:	HMS174(DE3); JM105; HB101;
	BL21(DE3)pLysE
Growth conditions:	Ap^{100}, Ap^{20}, Ap^{10}
	37°C, 30°C,
Culture conditions:	Standing, shaking
Media:	LB, LA, M9ZB
DNA extraction:	Variations of caesium chloride method,
	Nucleobond^{R} AX Kit PC 100 (Machery-Nagel, Germany),
	boiling method, and combinations of above.
Phenol treatment:	To remove nicked DNA and/or protein.

3.2.3 Induction of HMS174(DE3)(pET-12a-Bact-KY)

Once the desired plasmid has been obtained, the cloned DNA can be induced in HMS174(DE3) if the plasmid is stable. Therefore, immediately before induction each culture was titered to determine the fraction of cells carrying the plasmid. This was done according to a method previously described (Novagen Technical Bulletin, 1992). Cultures were plated on four different plates differing in the composition of the top agar. For induction experiments, the top agar was made 1 mM in IPTG at the time of plating. The format of the assay was as described below (Table 4).

Table 4. Plate assay to determine stability of pET-12a clones.

Cell dilution	Medium	Expected growth of:
(10^5 dilution)	IPTG + Ap	mutants which have retained plasmid but lost ability to express target DNA
	IPTG alone	only cells that have lost the plasmid
(2.10^6 dilution)	Ap	only cells with plasmid
	nothing	all viable cells

After the plate stability assay, three induction experiments were done. Initially, total cell lysates were analysed and thereafter the periplasmic fraction prepared by osmotic shock. 20 ml cultures of four HMS174(DE3)(pET-12a-Bact-KY) clones were grown in M9ZB Ap^{20} medium (Studier *et al.*, 1987) to log phase (~10^8 cells/ml), when the culture reached OD_{600} of 0.6-1 and then brought to 0.4 mM IPTG (time = 0). Absorbance readings (A_{600}) were taken at 0, 1, 2, and 3 h post-induction, and 1 ml culture samples taken at each time interval. Total cell contents were first analysed. Cells were harvested by centrifugation, placed in sample buffer (50 mM Tris-HCl (pH 6.8); 2 mM Na_3EDTA; 1% (w/v) SDS; 1% w/v) mercaptoethanol; 8%

(v/v) glycerol; and 0.025% (w/v) bromophenol blue), heated for 2-3 min in a boiling water bath and 50 µl samples applied directly to 20% SDS-polyacrylamide gels with a 6.7% stacking gel (Appendix; Laemmli, 1970). Gels were electrophoresed for 1.5 h at 100 V constant voltage until the bromophenol blue marker reached the end of the resolving gel and silver-stained by a method using copper sulphate enhancement according to Oakley *et al.* (1983).

Synthetic peptides used as molecular weight markers were a peptide of 986 Da (marker 1): H-Leu-Ile-Asp-Ser-D-Ala-Gln-Glu-Ile-Val-OH (9 amino acids) and a peptide of 1530 Da (marker 2): H-Asp-Ala-Glu-Asn-Leu-Ile-Asp-Ser-D-Trp-Gln-Glu-Ile-Val-NH4Ac- (13 amino acids; [D-Trp22]pHLRF, residues 14-26). The unprocessed and processed forms of the KY peptide have predicted molecular masses of 3454 Da and 1254 Da respectively.

A fractionation procedure for periplasmic proteins by osmotic shock was used, adapted from a method by Neu and Heppel (1965) and Kochland and Botstein (1980). Three fractions were isolated: the untreated control, the periplasmic fraction (supernatant), and the cytoplasmic and membrane bound fraction.

Cells were harvested by centrifugation for 2 min (14000 rpm) and the cell pellet transferred to ice, followed by addition of 150 µl cold sucrose solution, and 5 µl of EDTA, 0.5 M (pH 8), and kept on ice for 10 min. After microfugation for 5 min at 4°C, the sample was resuspended in 100 µl of cold dH_20. After 10 min on ice, the sample was microfuged again for 5 min, and the supernatant collected, this being the periplasmic fraction. The pellet, resuspended in 100 µl water, was the cytoplasmic and membrane-bound fraction. Sample buffer (as above) was added to each tube, and these were heated to 90°C for $2^1/_2$ mins before loading 50 µl volumes onto SDS-PAGE gels (20% acrylamide).

3.3 RESULTS AND DISCUSSION

3.3.1 Design of the pET-12a-Bact-KY construct

The folded mRNA of the transcript of pET-12a-Bact-KY was found to have a predicted free energy of -15.1 kCal/mol, which is acceptable for translation (Fig. 3.3).

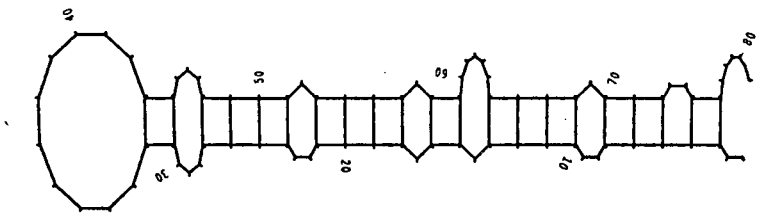


Fig. 3.3 Predicted secondary structure of the mRNA transcript of the pET-12a-Bact-KY fusion construct and its stability by free energy determination (Fold; Squiggles, GCG).

Pepplot plots measurements of protein secondary structure and hydrophilicity, showing several common measures of protein secondary structure on one coordinated plot (Fig. 3.4A). Dashed lines are alpha structures and solid lines, beta. Black is used for turns and hydropathy. The black curve is the Kyte and Doolittle hydropathy measure (Kyte and Doolittle, 1982). The line is in the upper part of the frame indicating hydrophobic regions and in the lower, hydrophilic. The Goldman, Engelman and Steitz (GES) curve (dotted line) is a similar program (reviewed in Engelman *et al.*, 1986).

A related program, (SAGA!!) plots the total positive and negative charges and the net charges of a protein as a function of pH (Fig. 3.4B).

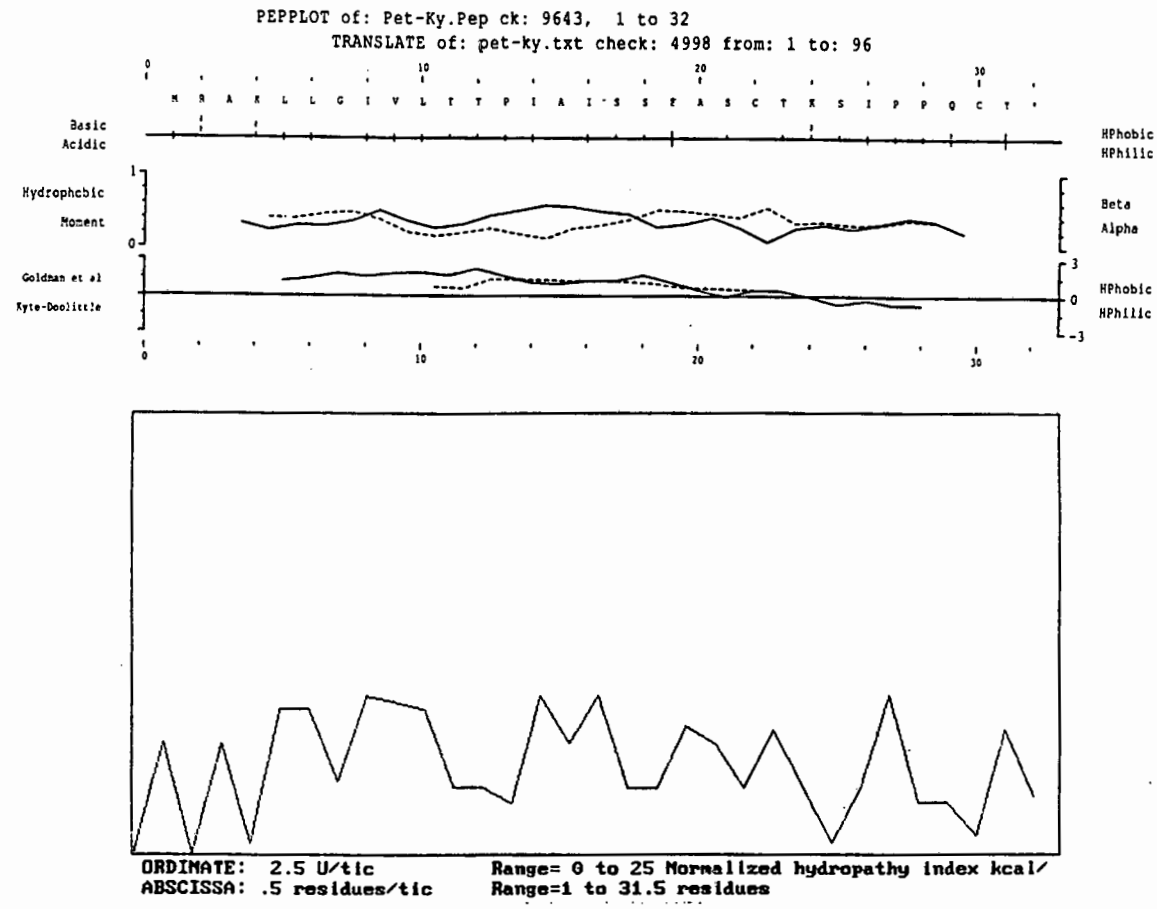


Fig. 3.4 Hydropathy plots of pET-12a-Bact-KY fusion peptide with the programs: [A] Pepplot (GCG); [B] SAGA!! (D. Maeder)

The first two highly charged residues of the peptide are glutamic acid and lysine, of the N domain of the OmpT leader. Of the first residues after the signal sequence splice junction Ser-Cys-Thr-Lys-Ser-, Lys has a high pK' value, the pH at which the acid is half-ionised. This is evident on the isoelectric scans (Fig. 3.5). As this residue is the P_1 residue of the active site of the inhibitor, no change here could be made to eliminate the high charge in this region.

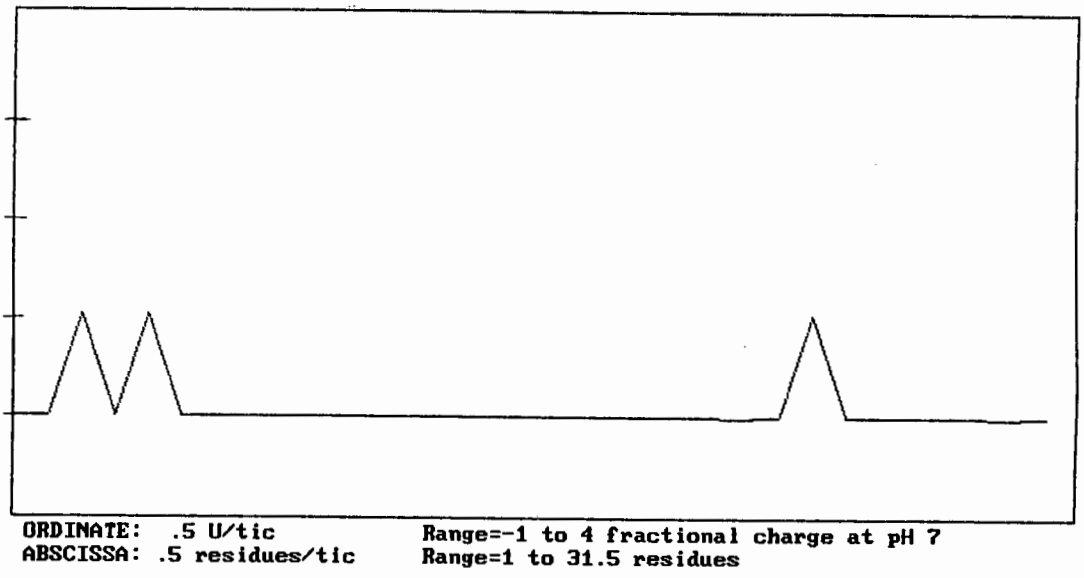


Fig. 3.5 Isoelectric profile of the OmpT-Bact-KY fusion peptide (SAGA!!; D.L Maeder; window: 1 residue). The two amino acids of the N-domain of the OmpT leader, Glu and Lys are of high charge, and Lys is again present as the P_1 residue of the active site of the KY peptide.

3.3.2 Lethality of KY in different *E. coli* strains

Gel purification of linearised pET-12a vector cut with *Spe*I was effective as no parental colonies were obtained when the DNA was used in test transformations. Colonies obtained from transformation of annealing mixes indicated successful cloning of the fragment. The cloning of the oligonucleotide was first indicated to have been successful from ampicillin-resistant colonies obtained in a controlled transformation experiment in which an equivalent quantity (62 ng) of prepared pET-12a vector (*Sal*I-cut, MBN-blunted and gel-purified) yielded zero colonies. Thus, no ligase being present in the annealing mixes, any cells able to grow on the Ap^{100} plates were expected to have arisen from transformation with vector having been able to recircularise after annealing of oligonucleotide into the cut ends, and repair of the gap.

Dot-blots probed with the oligonucleotide used in the cloning yielded false positives, as found after restriction analysis of maxiprepared plasmid DNA. This

was most likely due to the complementarity of the 15 bp left and right regions of this oligonucleotide with the region around the *Sal*I site of the vector. If the oligonucleotide loops out over the length of the central non-homologous residues, 30 homologous base pairs of the flanking arms could hybridise with vector ss-DNA. To overcome this, blots were subsequently probed with the 5' to 3' strand of the Bact-KY 54-mer which lacks these two regions. Clones pET-12a-Bact-KY-2, -8, -9 and -20 were identified from the strongest positive signals on these blots (Fig. 3.6).

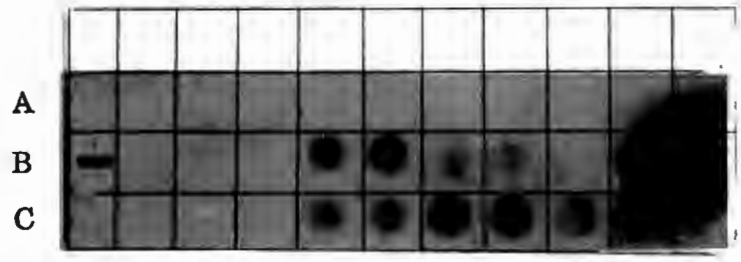


Fig. 3.6 Dot-blot hybridisation analysis of minipreparations of DNA from pET-12a-Bact-KY constructs in HMS174(DE3) probed with [γ-^{32}P]dATP 5' end-labelled Bact-KY oligonucleotide.Positive control (+): 50 ng Bact-KY oligonucleotide. Negative control (-): Miniprep DNA from HMS173(DE3)(pET-12a).

The presumptive clones in HMS174 grew poorly. Lysis of cells was apparent in liquid cultures grown overnight. After plasmid purification, the bands corresponding to hybrid pET-12a clones were faintly visible on agarose gels. The clones were transformed into JM105 and during maxipreparation of the DNA from these cells, loose pellets were obtained after the isopropanol spin and the clearing spin. Clearer gel resolution of the DNA was seen from JM105 than from HMS174 and the top two bands comigrated with pET-12a, but the lowest band (ccc DNA) was still not apparent in the presumptive clone preparations, the plasmid appearing to be nicked. Nicked DNA appeared to be obtained from Nucleobond AX preparations as well, as observed from agarose gels.

The annealing procedure and transformations were repeated in JM105 using newly prepared vector DNA and denaturation buffer. This yielded clones pET-12a-Bact-KY-9 and -16. However the cells still grew poorly. During maxipreparations of pET-12a-Bact-KY-9 and -16 from these cells, a cloudy streaming of the pellet was apparent upon spinning down overnight cultures, indicating cell lysis. Although the plasmid yield was improved in JM105, low yields were still obtained (32 μg from 1 l of culture). Other improvements were that the preparation was cleaner, the plasmid band was more clearly discernable on CsCl gradients, and the yields from the A_{260} scan accurately reflected the true plasmid yield as seen on agarose gels, as opposed to previous preparations in which artificially high readings were obtained due to degradation products. The sheared effect of the plasmid DNA was still seen on agarose gels and confirmation of the cloning event by restriction analysis was not possible.

Lysis of cells was most evident in shaking cultures at 37°C. A slight improvement occurred when the cells were grown at 30°C or as standing cultures (Studier, 1991). However, low yields of DNA were still obtained, viz. 21 μg and 7.3 μg from 1 l of culture vs 95 μg pET-12a/400 ml. Plasmid yields from JM105 were around 32 μg DNA from 1 l cultures.

Studier (pET technical bulletin) reports an improvement in growth of HMS174 strains in media with glucose such as M9ZB. This is thought to be due to the fact that T7 RNA polymerase production is less in this richer medium, although it is not supposed to be catabolite repressed. This was investigated and growth was found to improve, but plasmid obtained from these cultures was still not able to be sequenced.

Clones in *E. coli* HMS174(DE3) were transformed into strains JM105 and HB101 in an effort to obtain plasmid DNA of sufficient quantity and quality to sequence but this was not achieved even using a range of different growth and plasmid extraction conditions. In an attempt to confirm the clones in JM105 using restriction enzyme analysis of maxiprepared DNA on Hydrolink gels, cells were

grown as standing cultures. DNA was prepared from putative clones and restriction analysis attempted on 7.5% Hydrolink gels using *Bam*HI+*Xba*I and *Bam*HI+*Bgl*I. Whereas the excised *Bam*HI/*Xba*I (108 bp) and *Bam*H1/*Bgl*I (166 bp) fragments from pET-12a were observable, no bands were resolvable from the presumptive clones (Fig. 3.7). If the insert were present in these, fragments of 162 bp and 166 bp were expected respectively.

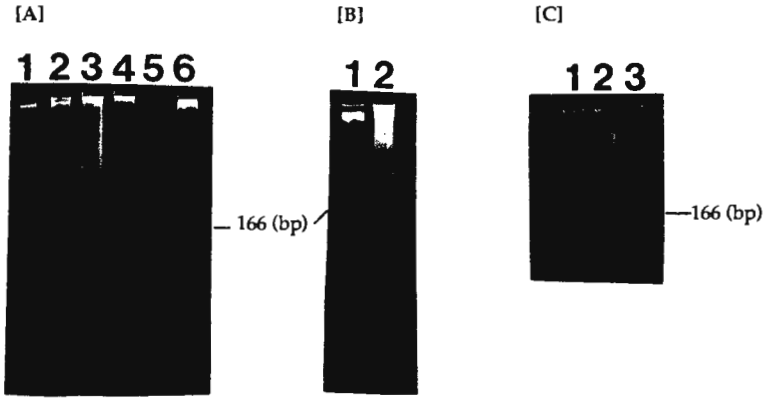


Fig. 3.7 Attempted restriction analysis of *Bam*HI/*Bgl*II digests of pET-12a-Bact-KY clones 7.5% Hydrolink gels. [A] Lanes 1 and 6, pET-12a control; 2, pET12a-Bact-KY9; 3, pET12a-Bact-KY6; 4, pET12a-BactKY16. [B] Lanes 1, pET-12a; 2, pET12a6. [C] Lanes 1and 2, pET12a-Bact-KY-6; 3, pET-12a.

3.3.3 Induction and attempted expression in *E. coli* HMS174(DE3)pLysE

Results of plate induction experiments (3.2.3) indicated nearly all the cells contained expressible target plasmids. However, this was an initial result on plates, in which lethality was not evident and this result gave no indication of later problems in aerated broth cultures. In the plate induction assay with unstable target plasmids, the fraction of cells that have lost the plasmid is reflected by an increase in the number of colonies on the IPTG plate and a decrease in the ampicillin plate as the cells lose the plasmid. An increase in the IPTG plate colony counts for BL21(DE3)pLysE cells was seen, but not a decrease in counts on ampicillin plates. The opposite trend to the above was observed from experiments with JM105 cells harbouring the construct. A problem with the interpretation of these results is that the lethality is not evident on plates in any event, but rather in the faster-growing, aerated broth cultures.

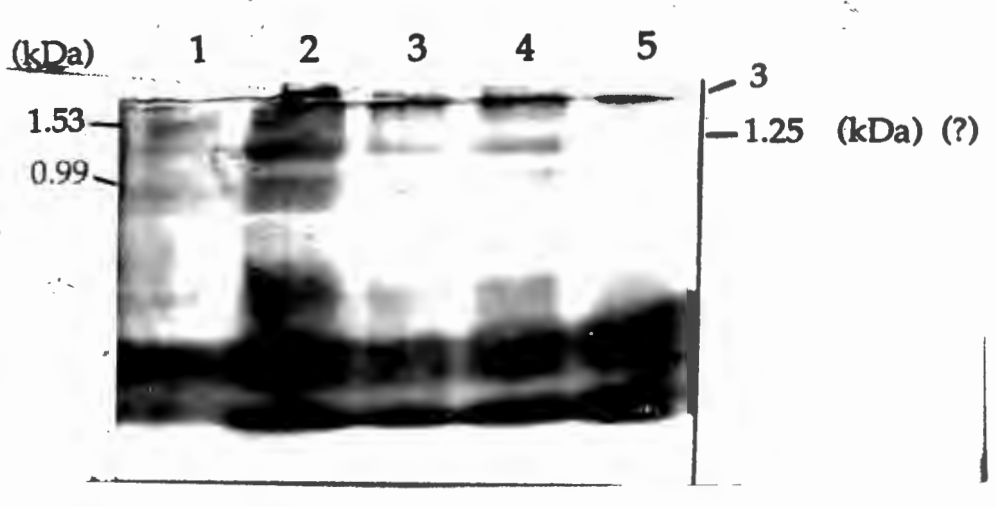


Fig. 3.8 PAGE gel analysis of total cell lysates from pET-12a-6 in HMS174(DE3)pLysE on 20% acrylamide. Lanes 1, Low molecular weight marker peptides; 2, post-induction; 4, post-induction; 5, before induction (time zero).

In analysing total cell lysates by PAGE after induction with IPTG, there were too many protein bands to be conclusive, although bands were present at the two relevant molecular weight ranges (Fig. 3.8).

3.3.4 Sequencing pET-12a-BactKY constructs

DNA was sequenced using a Sequenase™ kit supplied by United States Biochemical essentially according to the manufacturers' instructions. Attempted sequencing of templates revealed random priming, which may have been due to rearrangements or degraded plasmid DNA.

3.4 CONCLUSIONS

While pBR322 and most of its derivatives are relatively stable, problems of plasmid stability can arise when a gene whose product is toxic to the host cell is cloned in the plasmid. The expression level obtained can be such that the plasmid can be maintained but growth of the cell is impaired (Studier *et al.*, Novagen technical bulletin). In this case, cells lacking the plasmid can rapidly overgrow the culture. In cases where ampicillin is the selection antibiotic, this can happen when the culture is just becoming turbid, around 10^7 cells per ml. At this stage the β-lactamase enzyme reaches levels high enough to destroy the ampicillin in the medium. This point can be delayed by using 200 μg/ml ampicillin, but this does not keep the cells under selection all the way to saturation (stationary phase). The low yields of plasmid obtained from clones may have been due to this effect; cells losing the plasmid in the overnight broth cultures. This effect is also suggested by the improved growth of, and DNA yield from, cultures grown at ampicillin at 20 μg/ml.

Another complication reported by Studier *et al.* (Novagen technical bulletin) is that certain gene products are toxic in such a way that they, while having little effect on cells that are growing logarithmically, kill cells at saturation. As fewer cells retaining the plasmid survive up to saturation and as no ampicillin remains, cells that lack the plasmid overgrow the culture. This may explain the

effects seen in KY clones. (For this reason, as suggested (pET manual, Novagen) cells were grown and stored in M9ZB at -70°C.) It is interesting that in each case of plasmid instability investigated by Studier, substantial amounts of mRNA were found to accumulate in the cells and in some cases, considerable lysis of a culture was observed.

Another known problem is that the cloned DNA can be transcribed so actively that the cell becomes unable to continue dividing even if the mRNA or protein is innocuous (Studier, 1991). The level of T7 lysozyme provided by pLysE reduces the maximum transcription to a level at which induced cells can continue to grow, as long as the target gene products are not themselves toxic. The presence of T7 lysozyme can reduce the expression of cloned genes in BL21(DE3) with little if any gene product being found in uninduced cells.

Cells are thought to be able to tolerate high levels of lysozyme because of the enzyme's inability to penetrate the cell membrane to reach the peptidoglycan layer which is its substrate (Studier *et al.*, pET manual, Novagen). However treatments that disrupt the inner membrane but do not normally cause lysis, such as chloroform or mild detergents, induce rapid lysis of cells that contain even small amounts of T7 lysozyme (Moffat and Studier, 1987). Thus, if the expressed gene product were causing membrane destabilisation this could result in leakage of lysozyme through to the peptidoglycan layer and consequent lysis of BL21(DE3)pLysE strains harbouring a pET-12a-Bact-KY construct. Furthermore, the SOS response induced by certain proteins or peptides in BL21 cells, may cause lysis (pET manual, Novagen). If so, this should not occur in HMS174(DE3)pLysE cells. In accordance with this proposal, it has been noted that in some cases a cloned gene is less toxic in HMS174 derivatives of these strains (Studier, 1991). This is thought to be due to prophage induction from a small amount of a target gene product, which process requires recA, present in BL21 but not in HMS174.

It is not always possible to introduce any export signal sequence upstream of a gene and guarantee efficient export of the desired gene (Kadonago *et al.*, 1984; Ohsuye *et al.* 1983; Summers, 1989a). Export of a foreign translation product may

be affected if the recombinant protein contains sequences that are incompatible with and inconducive to tthe functioning of the cellular export apparatus (Stader and Silhavy, 1990). Attaching a signal sequence to such proteins can lead to death of the host cell. For example, fusion of ß-lactamase and proinsulin has proved to be toxic to *E. coli*, causing cell lysis (Brosius, 1984). It was suggested that in this case, the proteins may have become anchored in the cytoplasmic membrane, thus blocking export completely.

The charge of residues following the signal sequence is a potentially important feature. Summers (1988) found t-isomerase from chicken could not be secreted or processed when fused to the ß-lactamase signal peptide unless the Arg at position 3 of the mature protein was replaced. This was proposed to have been due to the extreme pK' of Arg. This problem was overcome by ion-pairing, by insertion of 2 Glu residues in front of Arg at the signal peptide junction. This lead to the question whether KY has amino acids with extreme pK' values at the N-terminal region. Similarly, the N-terminus of prochymosin contains a high proportion of charged hydrophilic/hydrophobic residues, viz. 2 Arg, 2 Glu, and 4 Lys residues in its first 20 amino acids. This abnormal distribution of charged residues is proposed to potentially disrupt the normal association of the N-terminus of the polypeptide with the secretory machinery (Little *et al.*, 1989). The high charge of the Lys residue of the KY peptide (Fig. 3.5) may cause problems in its transport through the membrane in the same way as these recombinant proteins.

Deng and coworkers (1990) proposed that one of the factors contributing to lethality in a bovine pancreatic phospholipase A2 construct could be the signal peptide, which was not cleaved in some of the expressed proteins. They proposed that the signal peptide in the fused protein might insert into the inner membrane of bacteria and affect the normal function of the membrane. This may have occurred with the KY peptide. Teixeira (1992), expressing *Erythrina* trypsin inhibitor (ETI) with OmpT found that most of the expressed protein with the signal sequence attached remained in the cytoplasm, as the cell's export system was unable to keep up with the level of expression.

Leemans and coworkers (1989) found that correct intracellular folding is required for a precursor protein to become processed and exported into the periplasm of *E. coli*. A polypeptide which by its own nature is insoluble in the cytoplasm of *E. coli*, is unlikely to be rescued from the cytoplasm by fusing it to an export signal peptide. They found that the export competence of human tumour necrosis factor (which has a temperature dependent intracellular solubility) correlated with the intracellular solubility of the protein. They postulated that the secretion proficiency of eukaryotic proteins, when fused to a prokaryotic export signal, depends on the ability of the mature protein to readily fold into a soluble conformation. Some proteins can be secreted into the periplasmic membrane when fused to a prokaryotic signal sequence, e.g. the human growth hormone (Hsiung *et al.*, 1986) and human interferon-alpha (Barbero *et al.*, 1986), whereas others, such as β-galactosidase or the chicken triose phosphate isomerase, remain unable to enter the export pathway (Moreno *et al.*, 1980, Kadonaga *et al.*, 1984). Hence, a signal sequence per se is insufficient to direct translocation from the cytoplasm and information within the mature protein seems to be required for correct sorting.

Overloading the export machinery may result from inefficient export of a foreign protein or be due to the protein being expressed at levels that exceed cellular capacity. Overloading also prevents the export of normal *E. coli* periplasmic proteins, precursor forms of normal and recombinant proteins accumulating in the cytoplasm, which can lead to cell death (Stader and Silhavy, 1990). This could be a slow process, and may be manifested to a greater extent in a rapidly growing, aerated culture as compared to a plate culture.

Strategies to prevent the problem of plasmid instability include the placing of a terminator upstream of the T7 promoter to prevent transcription from an upstream promoter region (Studier *et al.*, 1990), or eliminating nonessential regions of the plasmid that may contain weak promoters for *E. coli* RNA polymerase. This method was found to be effective in obtaining stability of a recombinant plasmid harbouring a plasmid transcriptional activator, which appeared to be lethal otherwise (Weiss *et al.*, 1991). Antisense promoters have

also been suggested to be useful for reducing basal expression of toxic target genes (Studier, 1990). Studier and coworkers (1990) found that an antisense promoter permitted the cloning and expression of certain genes previously impossible to clone, apparently by antagonising or neutralising the effects of promoters for *E. coli* RNA polymerase in the cloned DNA fragment.

The target protein itself may interfere with the cells metabolism or with the integrity of the cell. Cultures were found to grow better at a decreased growth temperature (30°C). Leemans and coworkers (1989), working with recombinant human tumour necrosis factor in *E. coli*, hypothesised that at low temperatures the synthesis rate of the protein was sufficiently slow to allow the export machinery to lead the molecule out of the cytoplasm. This theory is in line with the observation that secretion defects can often be overcome by slowing the rate of protein synthesis. Antibiotics, which do likewise are known to be suppressors of general export defects as are mutations in the protein synthetic apparatus. This effect may also explain lethality being most evident in the rapidly-growing aerated broth cultures as compared with plate cultures.

There is a small possibility that the ss oligonucleotide encoding Bact-KY may have annealled into another site in the vector other than the target site. In general, the larger the size of the mutation to be constructed, the lower the efficiency of oligonucleotide-mediated mutagenesis. This inefficiency stems from two sources. Firstly the ability of the mutagenic oligonucleotide to form stable hybrids with two separate sequences on the template DNA decreases as a function of the distance between the two sequences (Mandecki, 1986). Secondly, there is an increased probability that the oligonucleotide will hybridise to incorrect sequences on the template DNA. The consequence of this is that the frequency of regions of DNA carrying alterations at ectopic sites is greatly increased. It is therefore recommended to determine the complete sequence of the mutagenised DNA and not just the region that is the planned target. As the full sequence for pET-12a was not available from the suppliers, it was not possible to check for regions of homology between other sites in pET-12a vector and the oligonucleotide.

Thus, in the design of an expression construct, while a number of factors can be taken into consideration from principles known, the design process and choice of expression strategy can be open to experimental trial and error. Such strategies need to be formulated for each construct, each being unique, and outcomes are not always predictable (Texeira, 1992; Das, 1990). Difficult factors in the bacterial expression of KY included the smallness of the peptide. This made it difficult to screen for the DNA sequence and to detect the translation product on gels. Its lack of immunogenicity is another hurdle to overcome for immunodetection methods. Methods to overcome difficulties in expression of a protein showing toxicity include switching to a different expression organism, targetting the protein to a cell compartment in which its toxicity is not manifested, and regulating the expression of the protein with a regulatable promoter (Brown and Campbell, 1993). A construct for expression in the pMAL system (Maina *et al.*, 1988) was designed, but it is unsure whether this would also result in lethality. The bacterial KY construct was mainly intended for antibody production. The small size of the peptide did forewarn of problems here- low immunogenicity. A PCR-based method for screening for the KY DNA sequence in plants was concurrently developed (see Chapter 4), partially overcoming the requirement of antibody. For detection of the translation product, *in vitro* assays with synthetic substrate (Sunde, 1989; Maeder *et al.*, 1992) to determine activity of the recombinant peptide itself could be done instead of detection with antibody, and/or bioassays on transformed, regenerated plant material.

CHAPTER FOUR

A STRATEGY FOR THE EXPRESSION OF KY WITH A PLANT LEADER SEQUENCE IN TOBACCO

CONTENTS

CHAPTER FOUR

EXPRESSION STRATEGY FOR KY IN TOBACCO

4.1 INTRODUCTION

Utilisation of the synthetic KY peptide as a natural biocontrol agent against insect pests required the design and production of expression constructs for plants. To this end, both a construct with a leader sequence and a construct for direct expression without a leader sequence were designed for introduction into tobacco.

4.1.1 The GUS gene fusion system and the plant expression vector pBI121 for the expression of foreign genes in plants

Much work has been done on the development of plasmids for the expression of heterologous proteins in plants. The natural genetic engineering abilities of Agrobacterium have been harnessed to enable production of recombinant plants. *Agrobacterium tumefaciens*, a pathogenic soil-borne bacterium, induces crown-gall formation in wounded dicot plants, this ability being conferred by its large tumour-inducing plasmid, Ti. Transfer of a region of this plasmid, T-DNA, from the bacterial cells to the infected cells and integration into the nuclear genome of the plant initiates tumour formation.

The *vir* region of the Ti plasmid encodes proteins which are responsible firstly for the receipt of and response to phenolic signals released by wounded plant cells (Bolton *et al.*, 1986), and thereafter the production of a single-stranded region of DNA from the T-DNA, the T-strand, and its transfer to the plant cell. The T-DNA encodes biosynthetic enzymes for both the production of the plant hormones auxin and cytokinin- (oncogenic genes, resulting in the tumorous phenotype) and for production of specific opines, which are catabolised by Agrobacterium.

The octopine-type Ti plasmid contains a regulon of seven co-regulated operons, *vir*A-*vir*G. The products of these genes work in concert in detection of the wound

signal, activation of subsidiary *vir* genes, production and stabilisation of the T-strand from T-DNA, and transfer functions, including guidance of the T-strand to the nucleus and assistance in integration into plant nuclear DNA and tumour induction itself (Ream, 1989; Hooykaas, 1989; Zambryski, 1992).

These functions are harnessed for the production of recombinant plants carrying a gene of interest, by the use of *A. tumefaciens* strains in which the tumour-inducing genes have been deleted from the T-DNA. These genes are not essential for the transfer of the T-strand, and have been replaced by the DNA to be transferred. Such plasmids are termed 'disarmed'. The 25-bp left and right border repeat sequences flanking the T-DNA are employed in the transfer of the T-strand to plant cells and integration into the nuclear genome. This occurs by a mechanism analogous to conjugal transfer of bacterial plasmids.

For Agrobacterium-mediated transformation of plants for the expression of KY, a binary vector, pBI121 was chosen. This plasmid is one of the components of the GUS gene fusion system developed by Jefferson *et al.* (1986). Binary vectors contain selectable markers functional in plant cells, and cloning sites for the insertion of foreign DNA, flanked by 25-bp T-DNA border repeats. The broad host range origin of replication and the selectable markers are functional in *E. coli* and *A. tumefaciens* such that they can be selected for and maintained in both these strains. The binary vector pBIN19 (Bevan 1984) carries the 25-bp T-DNA border repeats, the neomycin phosphotransferase gene (*npt*II) as selectable marker, and multiple unique cloning sites for insertion of sequence for transfer. Unlike the co-integrate system, no recombinational steps are required for integration of the T-DNA region into the Ti plasmid, the *trans*-acting functions of the *vir* region on a resident disarmed Ti plasmid being utilised to transfer the modified T-DNA region from the plasmid maintained in *A. tumefaciens* to plant cells.

The GUS gene fusion system is designed to allow for construction of chimeric beta-glucuronidase genes and sensitive assays of the gene product, based on the *E. coli uid*A gene encoding beta-glucuronidase. The GUS gene was initially developed as a gene fusion marker in *E. coli* and in *Caenorhabditis elegans*, but

has more recently been used extensively for monitoring chimeric gene expression in plants. The pBI121 intermediate vector used in this study contains a neomycin phosphotransferase gene (*npt*II) as a selectable marker, which confers kanamycin resistance to transformed plants and is under the control of the constitutive nopaline synthase (NOS) promoter (Bevan *et al.*, 1983a; Bevan, 1984; Fig. 4.1). pBI121 contains the constitutive promoter from cauliflower mosaic virus (CaMV) upstream of unique *Xba*I, *Bam*HI and *Sma*I cloning sites.

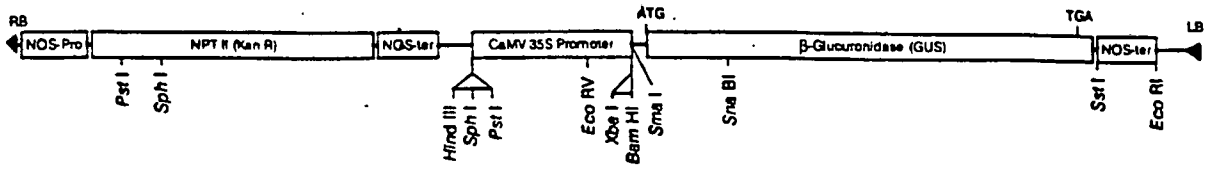


Fig. 4.1 Schematic representation of the T-DNA region of the pBI121 binary vector, showing the unique polylinker cloning sites *Xba*I, *Bam*HI and *Sma*I, between the CaMV 35S promoter and the GUS gene. The GUS gene is followed by the terminator (NOS-ter) and polyadenylation site of the nopaline synthase gene (Clontech manual).

4.1.2 Design of the plant leader sequence fusion protein Tomwipig-Plant-KY

Alternative strategies for the expression of KY in plants were formulated in order to optimise activity and expression levels. Protein and DNA features influencing activity, expression and gene regulation of proteinase inhibitors in plants, both at the transcriptional and translational levels were considered from the literature (discussed in 1.3.3) and EMBL and Pir (Protein Information Resource) database searches.

Hilder *et al.* (1989a) reported that since the available evidence from legumes indicated that proteinase inhibitors are primarily cytosolic proteins, no signal sequence should be necessary. However, a number of authors have noted that different plant proteinase inhibitors possess signal sequences and/or reported their localisation in either the vacuole or vacuole-derived protein bodies (1.4.3). To further investigate the prevalence of signal sequences in the protease inhibitors of plants, the GenEMBL database was searched using Stringsearch (GCG), investigating the Bowman-Birk inhibitors of legumes in particular. The literature and the Pir and GenEMBL databases were searched with the further aim of examining DNA sequences controlling the expression of plant proteinase inhibitors, especially wound inducibility, and protein motifs involved in targeting of PIs to specific locations in the cell.

Of particular interest were motifs within mature proteins involved in their targeting. Sebastiani and co-workers (1991), using the computer programs Gap, Find, Peptidestructure and Plotstructure (Devereau *et al.*, 1984), conducted amino acid sequence comparisons of plant proteins sorted to protein bodies and plant vacuoles, in order to identify shared topological equivalents involved in sorting. Comparisons made between 34 members of different plant protein families and the soybean α subunit of β-conglycinin revealed that while overall sequence similarity between different members of protein families sorted to these organelles is low, there are certain amino acid sequences that are conserved. Sequence conservation is considered not to be as important at the protein level as overall structural (topogenic) conservation. Sebastiani and coworkers identified a number of short conserved sequences from these seed storage proteins, plant proteinase inhibitors and lectins (Fig. 4.2).

```
PROTEIN                AA POSITION          LOCAL     PRIMARY     SEQUENCE          AA POSITION

GROUP I (7S storage proteins)
                                      A                                  B                          C
                                      ----                               ---------                  -------
α subunit, soybean          192  IRVLQRFNQ...RSPQLQNLRDYR.I.........LEFNSKPNTLL.LPNHADADYLI 235 > 426 GALLLPHFNSKAIVIL
β subunit, soybean           30  IRLLQRFNK...RSQQLENLRDYR.I.........VQFQSKPNTIL.LPHHADADFLL  73 > 259 GALLLPHFNSKAIVIL
Phaseolin, bean              33  IRVLQRFDQ...QSKRLQNLEDYR.L.........VEFRSKPETLL.LPQQADAELLL  76 > 254 GALFVPHYYSKAIVIL
Vicilin, pea                 27  IRLLQKFDQ...RSKIFENLQNYR.L.........LEYKSKPHTIF.LPQHTDADYIL  70 > 269 GSLLLPHYNSRAIVIV
Convicilin, pea             145  IRRLQRFDK...RSDLFENLQNYR.L.........VEYRAKPHTIF.LPQHIDADLIL 188 > 385 GALMLPHYNSRAIVVL
Canavalin, Canavalia         41  LRLLQRFNE...DTEKLENLRDYR.V.........LEYCSKPNTLL.LPHHSDSDLLV  84 > 265 GALFVPHYNSRATVIL
Vicilin, Vicia               27  IRLLQKFDQ...RSKLLENLQNYR.L.........LEYKSKPHTIF.LPQQTDADFIL  70 > 269 GSLLLPHYNSRAIVIV
Vicilin, cotton             175  FRVLQRFAS...RHPILRGINEFR.L........SILE..ANPNT.FVLPHHCDAEKIY 118 > 403 GSIFVPHYNSKATFVV

GROUP II (11S storage proteins)
                                      A                                  B                          C
                                      ----                               ---------                  -------
Legumin, pea                 50  RATLQRNAL...RRPYYSNAPQEIFI..QQGNGYFGMVFPGCPETF.EEPQESEQGEGR 102 > 368 NAMFVPHYNLNANSII
GY2 glycinin, soybean        50  RCTLNRNAL...RRPSYTNGPQEIYI..QQGNGIFGMIFPGCPSTY.QEPQESQQR.GR 101 > 339 NAMFVPHYTLNANSII
GY4 glycinin, soybean        49  KLTLNRNGL...HLPSYSPYPRMIII.AQQGKGALGVAIPGCPETF.EEPQEQSNRRGS 102 > 412 NGIYSPHWNLNANSVI
Legumin, Vicia                5  FDRLNQCR-49AA-PSYSPSPQLIYII..QGKGVIGLTLPGCPQTY.QEPRSSQSRQGS 103 > 231 NGIYAPHWNINANSLL
Globulin, oak                23  FDRLQAFE-41AA-PQYYHNAPGLVYIL.QGRGFTGLTFPGCPATFQQQFQPFDQAR.F 114 > 350 NAILSPYWNINAHSVM
Glutelin, rice               23  FDRLQAFE-41AA-SIHKIS.PGVVYII.QGRGSMGLTFPGCPATYQQQFQQFSSQGQS 114 > 335 NAILSPFWNVNAHSLV
C134, cotton                 68  RHKIQRKGL....LPSFTSAPMLFYV..EQGEGIHGAVFPGCPETYQSQSQQNIQDRPQ 221 > 369 NAIYAPHWNMNAHSIV

GROUP III (Protease inhibitors)
                                      A                                  B
                                      ----                               ---------
Kunitz Tryp Inh, Erythrina    7  GEVVQNGGTYYL.LPQVWAQGGGVQLAK......TG..EETCPLTVVQSPNELSDGKPI  56 > *
Kunitz Tryp Inh, wingedbean   8  GELVRNGGTYYL.LPDRWALGGGIE.AAA.....TG.TE.TCPLTVVRSPNEVSVGEPL  57 > *
Kunitz Tryp Inh, soybean      7  GNPLENGGTYYI.LSDITAFGG.IR.AAP.....TG..NERCPLTVVQSRNELDKGIGL  55 > *
Browman-Birk Inh, alfalfa     1                                    TTACCNF.CPCTRSIPPQCRCTDIGE  26 > *
Protease Inh, potato          1                                KACIRECGNLGFGICPRSEG.SPENPICTNCC  31 > *
Protease Inh, tomato          1                                KACTRECGNLGFGICPRSEG.SPNLNPICINC  31 > *

                                      A
                                      ----
CPY, yeast                    1   ISLQRPLGLDKDVLLQAAEKFGLDLDLDHLLKELDSNVLQAWAQIEHLYPNQVMSLET   58 >
```

Fig. 4.2 Sequence alignment of the partial amino acid sequence of the soybean α-subunit of β-conglycinin relative to the 7S and 11S storage proteins, and protease inhibitors sorted to protein bodies and plant vacuoles. The amino acid position is relative to the N terminus of the precursor protein after the signal peptide is cleaved, for those precursors which have a cleavable signal (Sebastiani *et al.*, 1991).

Table 3. Results of GenEMBL Search for Evidence of Signal Sequence among Pis[a]

LOCUS	ACCESSION	SOURCE	COMMENTS	REFERENCES
POTINHWI	M17108	potato	WI, pre-, (Ala-Arg)	Cleveland *et al.*, 1987
POTPI2G	X04118	potato	pre-	Keil, 1986
POTPIIIRI	X03778	potato	PI IIcDNA pre-	Sanchez-Serrano *et al.*, 1986
POTPPIIIK	M29965	potato	WI IIK pre-	Palm *et al.*, 1990
STAPI	X53470	potato	aspartic, pre-, 32	Strukelj *et al.*, 1990
RICCPI	J03469	rice	seed specific	Abe *et al.*, 1987
TOMPII	J05094	tomato fruit	dev. regl. wild pre-	Wingate *et al.*, 1989
TOMPROI	M16258	tomato	ethylene induced	Lincoln *et al.*, 1987
TOMWIPI	K03290	tomato leaf	10 bp pal WI, pre-pro, CTPP	Graham *et al.*, 1985
TOMWIPIG	M13938	tomato	WI, pre-pro, CTPP	Lee *et al.*, 1986
TOMWIPII	K03291	tomato	WI, pre-	Graham *et al.*, 1985
ST322R	X13180	30% homol. BBI	pre-,27; pro-,20	Stiekema *et al.*, 1988
BLYPAPI2	M15207	barley	Prob amylase/PI pre-	Mundy *et al.*, 1986
SOYCIIPI	K01967	soybean BBI	pre-,27 (Lys/Ser)	Hammond *et al.*, 1985
SOYCIIPIA	M20732	soybean	cDNA PI IV and C II	Joudrier *et al.*, 1987
SOYCIIPIB SOYCIIPii	M20733	soybean	PI-IV mRNA	Joudrier *et al.*, 1987
BOVPTIEP A		artif. BPTI	alk. phos sig. from *E. coli*	Marks *et al.*, 1986

[a]: List of Abbreviations

PI: proteinase inhibitor
pre-27: signal peptide of 27 amino acids
(T/T): trypsin/trypsin specificity
(T/C): trypsin/chymotrypsin specificity
(T/E): trypsin/elastase specificity
dev. regln: developmental regulation
cDNA: copy DNA sequence pre-:
pre-: prepeptide
pro-: pro-peptide
BBI: Bowman-Birk inhibitor
WI: wound-inducible
CTPP: C-terminal propeptide

The PIs investigated, from Leguminosae and Solanaceae, were found to have a sequence related to a conserved sequence designated region B of the 7S and 11S storage proteins of Leguminosae. This sequence maps on the surface of the tertiary structure of phaseolin and soybean Kunitz trypsin inhibitor as do the other homologous sequences identified, on proteins for which three-dimensional structures are known, consistent with their proposed role as sorting signals. Region B, (K-P-N/H-T-(X_3)-P), present as the sequence CPCTRSIPPQ (Fig. 4.2) in the Bowman-Birk inhibitor of alfalfa, is also present in the KY peptide from DE3 as the sequence SCTKSIPPQ. In the 11S storage proteins and PIs, the Cys residue, which replaces the Lys residue of the 7S storage proteins, is involved in the formation of an intramolecular disulphide bond.

It is therefore proposed that the DE3 inhibitor of *M. axillare* is naturally targeted to the vacuole *in planta* and that it would be possible for the KY peptide, provided with a leader sequence to enable transport through the endoplasmic reticulum, to be similarly targeted upon expression in a heterologous plant, due to possession of this topogenic sequence.

To investigate further sequences involved in targeting, signal sequences, (pre- and pro-sequences,) and vacuolar targeting motifs including C-terminal propeptides (CTPPs) were investigated. Plant trypsin inhibitors from tomato, potato, soybean, barley, poplar and rice were identified using GenEMBL Stringsearch (Table 3).

The tomato and potato PIs analysed are produced as prepropeptides with a signal peptide of usually 23-27 amino acids, processed at the splice junction Ala/Arg or Ala/Ser and sorted to the vacuole. The protein sequences located from the Pir database included those of three Bowman-Birk inhibitors from soybean. Protease inhibitor D-II precursor IV, a BBI (T/T) of 83 amino acids (accession PO1064) is an inhibitor which has no signal sequence but has a propeptide, MCILSFLK. (Joudrier *et al.*, 1987; Odani, 1978). Another BBI (accession P01055; T/C) of 71 amino acids has seven disulphide bridges (Odani and Ikenaka, 1972). Thirdly, a BBI of 83 amino acids (PI C-II precursor; accession IBB2$SoyBN) has two active

sites interacting with elastase and trypsin/chymotrypsin. This inhibitor has a propeptide from residues 1-7: MEFNLFK (Joudrier *et al.*, 1987; Hammond *et al.*, 1984; Hammond *et al.*, 1985; Odani and Ikenaka, 1977). Thus although none of these inhibitors had evidence of a signal peptide, they are apparently targeted by their propeptides. These have seven disulphide bridges. A fourth BBI from soybean (Soyciipi), identified from the GenEMBL search, has a signal sequence of 27 amino acids (Hammond *et al.*, 1987).

Vacuolar targeting signals may occur at the beginning, at the end or within the mature protein. From the literature (1.4.3) and GenEMBL database searches, it was evident that the protease inhibitors of tomato, potato and soybean are frequently found in the vacuole of the cell, to which they are directed by a signal sequence ahead of the protein itself. This arrangement would also be of advantage to a peptide such as KY in protecting such a small peptide from intracellular breakdown and in delivering the peptide into the cellular compartment where disulphide bridge formation is favoured.

The similarities between signal peptides of different proteins are in their general structural features and not sequence identity (Pugsley, 1993). Thus, a 'typical' signal peptide, of those identified from the GenEMBL database search, the 20 amino acid leader of the tomato wound-inducible proteinase inhibitor gene, Tomwipig, was chosen to be cloned as a fusion to the plant KY sequence. The amino acid sequence of the leader was reverse translated using plant codon usage (Konigsberg and Godson, 1987) and aligned with the Plant-KY sequence to form a DNA sequence designated Tomwipig-Plant-KY, encoding a 33-residue fusion protein. It was decided to use the signal sequence from the Tomwipig gene with its natural *in vivo* splice junction, Ala/Arg, as this would be processed correctly in tobacco, since this junction is common to most plant signal sequences, including those of tobacco .

4.1.3 Design of oligonucleotides to encode the plant leader construct Tomwipig-Plant-KY

It was not possible to clone a leader directly in-frame with the Plant-KY sequence in pSK-Plant-KY as the plant spacer sequence AACA is present in this construct upstream of the start codon (2.3.2), and no restriction sites were present with which to excise it. This required the plant signal sequence to be fused in frame with a new Plant-KY sequence. In order to clone the gene into pBI121 without creating an unwieldy GUS fusion, a pBI121ΔGUS construct was used from which the GUS gene had been deleted at the *Sma*I/*Sac*I sites, leaving the *Bam*HI/*Xba*I sites for cloning. This vector retained the NOS terminator and the polyA signal of GUS, important in enabling movement of the mRNA to the cytoplasm and in extending the half-life of the message (Gallie, 1993). The Plant-KY sequence was to be subcloned from pSK-Plant-KY into the plant expression vector pBI121ΔGUS as a blunt *Cla*I/*Bam*HI fragment.

The Tomwipig-Plant-KY construct was designed to be cloned with the oligonucleotide double-strand break repair cloning method of Mandecki (1986; Fig. 4.3A). As the host cell's fill-in reaction is accurate for only short stretches of ssDNA, around 60 nucleotides, the sequence was designed to be cloned using two oligonucleotides, the first encoding the first 15 amino acids of the Tomwipig signal sequence and the second, the 17th to the 20th amino acids of the signal, the 11 amino acids of the KY sequence and the terminator codon (Fig. 4.3B). Gap distances to be filled in after annealing of these oligonucleotides would be 49 bp and 51 bp respectively. In order to open the vector again after cloning in oligo I, a restriction site for cloning was searched for which would be present in neither the pBI121ΔGUS vector nor the cloned sequence. This information being unobtainable for the vector from the suppliers of the pBI plasmids, the pBluescript vector system was chosen to be used for the initial cloning steps of the oligonucleotides, having the advantage of a fully known sequence, multiple cloning sites, ease of cloning and sequencing, and the facility to flip the cassette in

site not to be otherwise present in the sequence to be cloned. Supplying the *Spe*I site at the 3' end of oligonucleotide I would enable reconstitution of the site, providing for the subsequent cloning of oligonucleotide II.

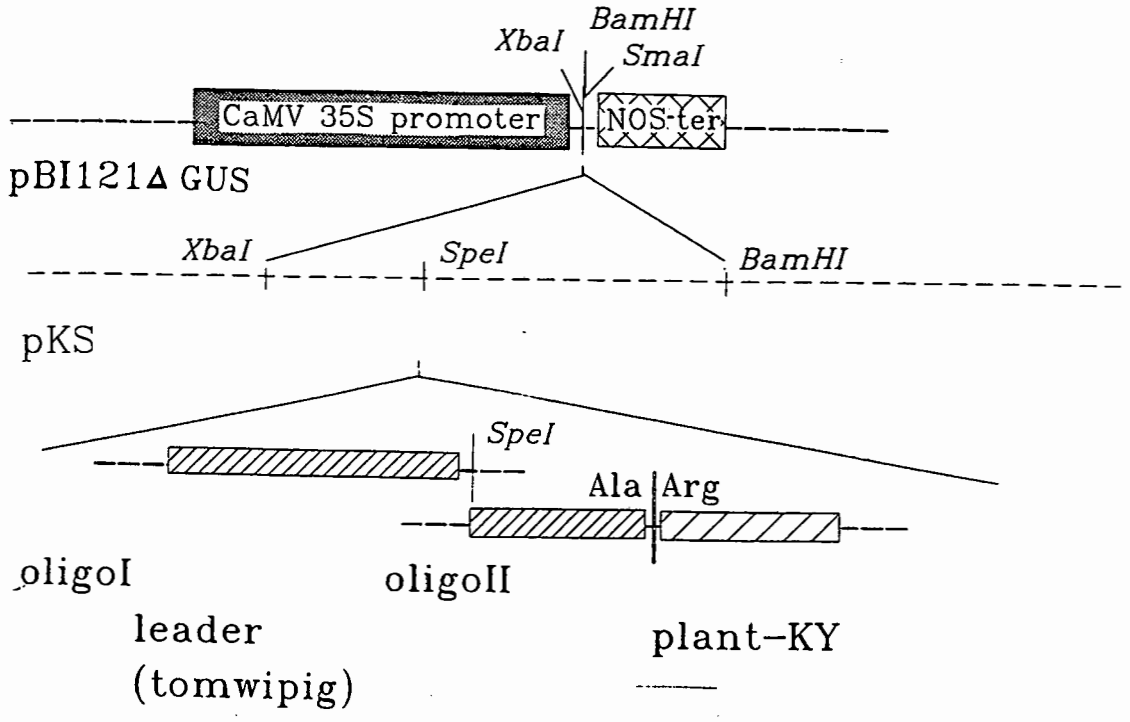


Fig. 4.3 **[A]** Schematic representation of the construct Tomwipig-Plant-KY designed to be cloned into pBI121ΔGUS utilising oligonucleotide-directed bridge-break repair. Cloning into the intermediate vector, pKS supplies the *Xba*I and *Bam*HI sites to the construct. The Tomwipig leader and KY peptide sequences are represented by boxes of narrow and broad stripes respectively. The splice junction of the Tomwipig leader at Ala/Arg is indicated with an arrow. The LHS and RHS 15-bp arms of oligomer I and the RHS arm of oligomer II are complementary to the *Spe*I site region of the pKS vector, enabling annealing into this site. The RHS arm of oligomer I regenerates the *Spe*I site upon bridge break cloning. The LHS arm of oligomer II is complementary to the 3' terminus of oligomer I upstream of the *Spe*I site.

After confirming the sequence of the Tomwipig-Plant KY construct, directional subcloning of this fragment into pBI121ΔGUS would be facilitated at the *Xba*I and

*Bam*HI sites. As the natural plant leader splice junction Ala/Arg of the tomwipig leader was used, upon removal of the signal peptide, Arg would become residue 1 of KY. This was predicted not to interfere in any way with activity (pers. comm., D. Botes).

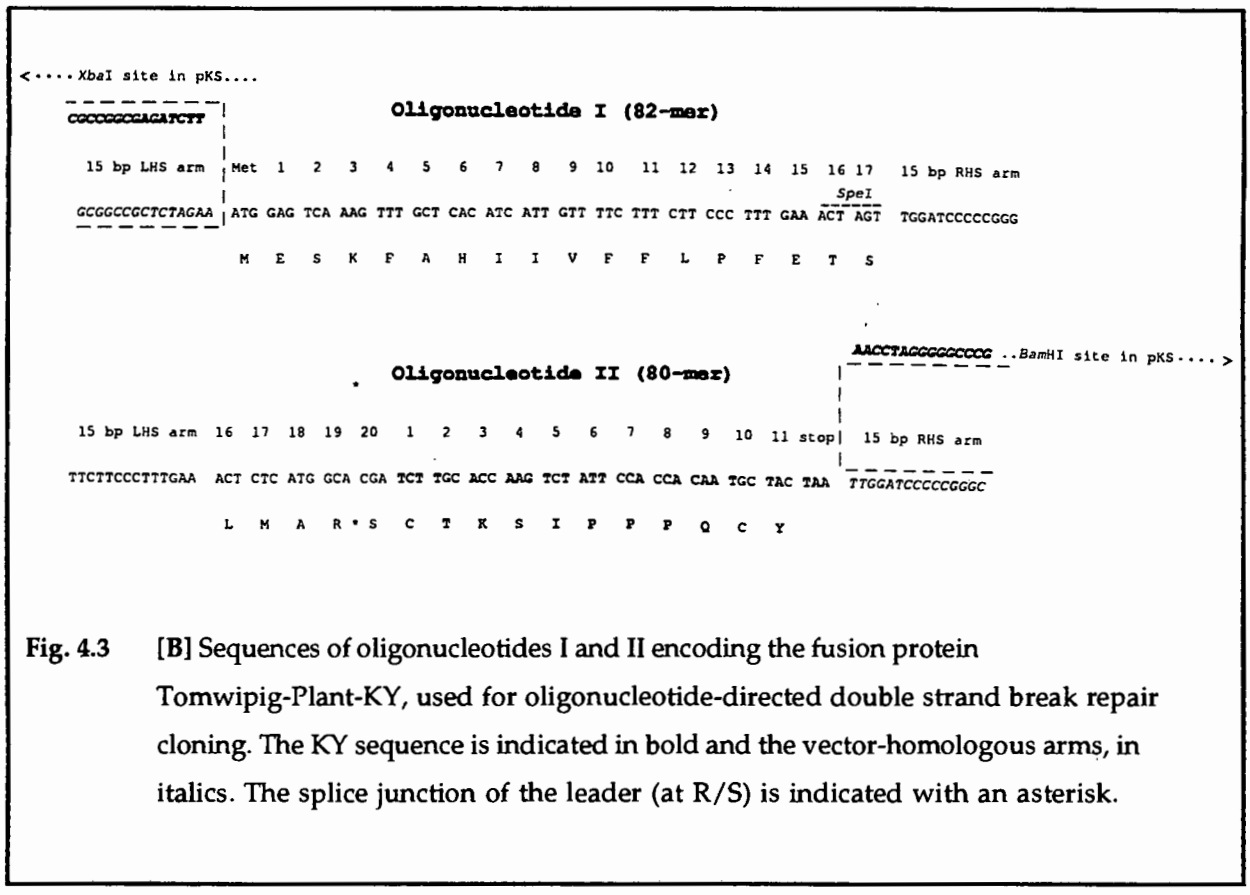


Fig. 4.3 [B] Sequences of oligonucleotides I and II encoding the fusion protein Tomwipig-Plant-KY, used for oligonucleotide-directed double strand break repair cloning. The KY sequence is indicated in bold and the vector-homologous arms, in italics. The splice junction of the leader (at R/S) is indicated with an asterisk.

To assess the probability of the oligonucleotide annealing into the pKS vector at an ectopic site, a site other than that of the target, a FASTA sequence comparison search was done with the full oligonucleotide sequence including the left and right 15-bp arms.

4.2 MATERIALS AND METHODS

4.2.1 Design of the fusion peptide Tomwipig-Plant-KY and encoding oligonucleotides

The hydrophilicity profile of the Tomwipig-Plant-KY fusion peptide was analysed using the program Peptidestructure (GCG), and plotted by the Hydropathy

program of the SAGA!! protein sequence analysis package (D.L. Maeder, University of Cape Town), using the Kyte-Doolittle data set and by Pepplot (GCG). Peptidestructure predictions include hydrophilicity calculated according to the algorithm of Hopp and Woods (1981) for which the window was set at 1.

4.2.2 Cloning by oligonucleotide-directed bridge-break repair into pKS

i. Purification of oligonucleotides

Oligonucleotides were purified firstly by PAGE on vertical nondenaturing 10% acrylamide gels, and secondly by HPLC. The oligomer was eluted from the gel essentially according to the method of Ausubel (1987). The gel was cut into small pieces, crushed and the DNA eluted overnight in twice the gel volume of elution buffer (0.5 M ammonium acetate, 1 mM EDTA and 0.1% SDS) and precipitated as described. The supernatant was filtered through siliconised glass wool packed into the tip of a pierced PCR tube. The DNA was then precipitated ammonium acetate, the pellet obtained rinsed with 70% ethanol, dried briefly in a Speedyvac and resuspended in the desired amount of TE buffer (10 mM Tris-Cl, pH 7.5, 1 mM EDTA).

HPLC purification of the oligonucleotides was undertaken using a Millipore Waters pump system with an anion exchange Mono-Q™ HR10/10 (10 cm X 10 mm) cartridge. The elution solvent was 10 mM NaOH using a LiCl gradient. The gradient was established using a Buffer A of 50 mM LiCl, 10 mM NaOH, pH 12 and Buffer B of 3 M LiCl, 10 mM NaOH. The column was equilibrated by washing with A for 20 min, B for 10 min and again with A for 20 min before loading the sample in a small volume of Buffer A (500 µl). The gradient conditions used for the elution were established with the following program at a flow rate of 2 ml/min.

Time (min)	A%	B%
0-5	100	0
5-25	95	5
25-225	95-50	5-50

The eluted oligonucleotide was precipitated with butanol at a ratio of 1:10, and then with 1:3 ethanol: acetone and ammonia-butanol as described (2.2.3). The sample was detritylated with 80% acetic acid over 1 h at room temperature, and reprecipitated with butanol before drying with a Speedyvac, and the O.D. read at A_{260}.

ii. Cloning of oligonucleotides

The Bluescript vector pKS was digested with *Spe*I, purified through 0.7% TAE preparative gels, recovered by the Geneclean method (BIO101, USA), and quantified by comparison with prepared standard DNA samples. Amounts of vector and oligonucleotide used were established by calculation of picomolar ends. For the optimal ratio of oligo: vector inferred from Mandecki (1986) 39 ng of prepared vector (20 fmol) and 545 ng of oligonucleotide (20 pmol; 1000X molar excess) were annealed together in a reaction volume of 30 µl, buffered with denaturation buffer (10 mM Tris-HCl pH 8.0, 10 mM KCl, 5 mM $MgCl_2$), and incubated at 100°C for 3 min in a boiling water bath. The reaction was allowed to cool at room temperature for 5 min before transfer to ice and addition of 100 µl of competent cells prepared by the calcium chloride (Sambrook *et al.*, 1989) or rubidium chloride (Mitage *et al.*, 1988) methods. Cells were transformed as described for these methods.

In order to establish a standard curve of numbers of transformants obtained against different oligonucleotide: vector ratios used (as in Mandecki, 1986), the quantity of vector used was maintained at 20 fmol against a range of oligonucleotide quantities -viz. 0.1; 0.5; 1; 5; 10; 20; 50 and 100 pmol. This would also account for any differences in concentrations of oligonucleotide calculated from O.D. readings and actual concentrations.

In order to narrow the search for transformants, scaled-down experiments were subsequently done around the optimal 20 pmol oligo I : 20 fmol vector ratio. In these, 20 pmol oligo I were annealed with a range of three vector quantities in order to account for any inaccuracy in estimating the concentration of vector DNA after gel purification and recovery by the Geneclean procedure, as these

preparations cannot be scanned at A_{260}. Initially, the vector was treated with Mung Bean nuclease to polish the ends but this was later found to be unnecessary. Cells used for transformations were *E. coli* JM105, JM109 and JM83, and these were plated on LA Xgal plates incorporating ampicillin (100 µg/ml).

4.2.3 The identification of recombinant clones in pBluescript

Blue-white selection and *Spe*I restriction analysis of miniprepared plasmid DNA were first used to screen for transformants of oligonucleotide I. Colonies were additionally screened by dot-blot hybridisation of minipreparations of DNA as described in 2.2.4. Oligonucleotide I was used as probe, as no other DNA carrying the Tomwipig signal sequence was available.

A method enabling PCR amplification of the MCS of pKS by T3 and T7 primers was subsequently developed to enable a more quantitative assessment of whether presumptive clones contained the 49-bp sequence of interest. Parameters were changed in optimising the PCR reaction from a starting test protocol for pSK. A decreased amount of primers and the longer sequencing primers, T3 and T7 of 26 bp each were used, (vs 17 bp), at a final concentration of 1 mM. Primers were adjusted to a concentration of 50 mM in dH_20 and stored at -20°C. Tween 20 was added to 0.05% of reaction buffer (a 0.5% stock added to the multimix at a ratio of 1:10). Some Taq enzymes act better with detergent, which prevents the natural tendency of the enzyme to aggregate. The annealing temperature was lowered to 50°C from 60°C, according to the Tm values of the primers (Primer program). A lowered dNTP concentration (1 µl of 50 mM stock) was used. The amount of target DNA was reduced by use of fewer cells, as this resulted in a clearer product with decreased background smearing on gels. Cell aggregation obscures the PCR product on gels and possibly hinders the reaction. Broth culture screening was found to be the cleanest. Keeping concentrations of DNA and dNTPs low avoided artefacts -false bands arising from mis-priming.

As OD readings do not always reflect the true amount of oligonucleotide in the sample, the primers were run on vertical 12% acrylamide gels to confirm that

their concentrations were within the expected range. They were adjusted to 50 mM stocks, aliquotted in dH_20 and stored at -20°C. TE was avoided as EDTA chelates Mg^{2+} ions, which are required by the Taq enzyme. The primers were of the following sequences:

T7: 3' GGGTTAAGCGGGATATCACTCAGCAT 5' (26-mer)
T3: 3' GGGAAATCACTCCCAATTAAGGCTCG 5' (26-mer)

To screen colonies by PCR, cells were inoculated with a toothpick into 100 µl of dH_20, boiled for 3 min in a water bath and snap-cooled on ice. Aliquots of 25 µl of cell suspension were added to the same volume of 2X PCR reaction mix in PCR tubes and overlaid with 40 µl autoclaved paraffin oil. PCR amplification reactions contained 1 mM of each primer (10 pmol) in a total reaction volume of 50 µl. Tween 20 was added to a final concentration of 0.05% in the multimix and 200 µM of a dNTP cocktail mix (ie. 50 µM of each nucleotide) was used (Appendix B, Table I). Taq™ enzyme and buffers used were supplied by Promega.

Reaction mixtures were subject to 25 rounds of amplification through a denaturation step of 92°C (20 sec), an annealing step of 50°C (30 sec) and an extension step of 72°C (300 sec). DNA amplification cycles were controlled by a custom-made programmable thermocycler (JDI model 8012). After amplification, samples were electrophoresed through 3% agarose gels incorporating 0.5% EtBr and photographed under UV light.

Recombinant plasmids identified by PCR screening were sequenced and analysed with the FASTA program (GCG) to confirm the presence of target DNA sequence. FASTA does a Pearson and Lipman sequence similarity search between a query sequence and the GenEMBL database (Pearson and Lipman, 1988).

4.3 RESULTS AND DISCUSSION

4.3.1 Design of fusion peptide and encoding oligonucleotides

The protein analysis program of GCG, Pepplot plots measurements of protein secondary structure and hydrophobicity, showing several common measures of protein secondary structure on one coordinated plot. The protein showed a characteristic hydrophobic core region flanked by hydrophilic N-terminal residues (Fig. 4.4).

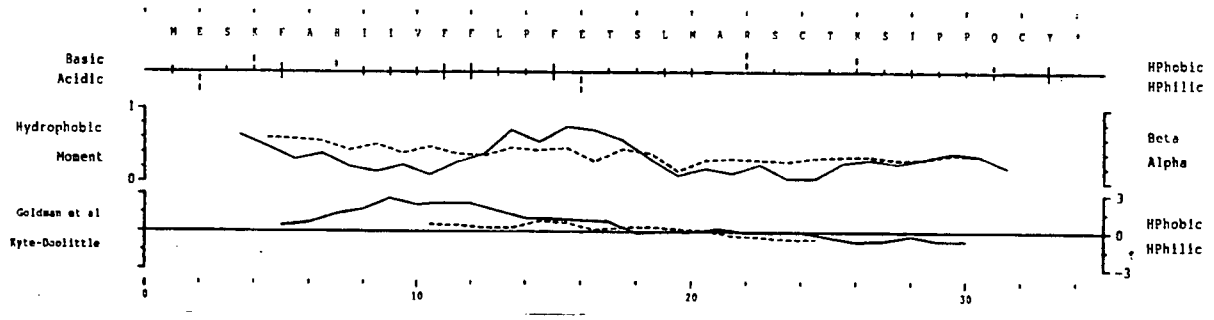
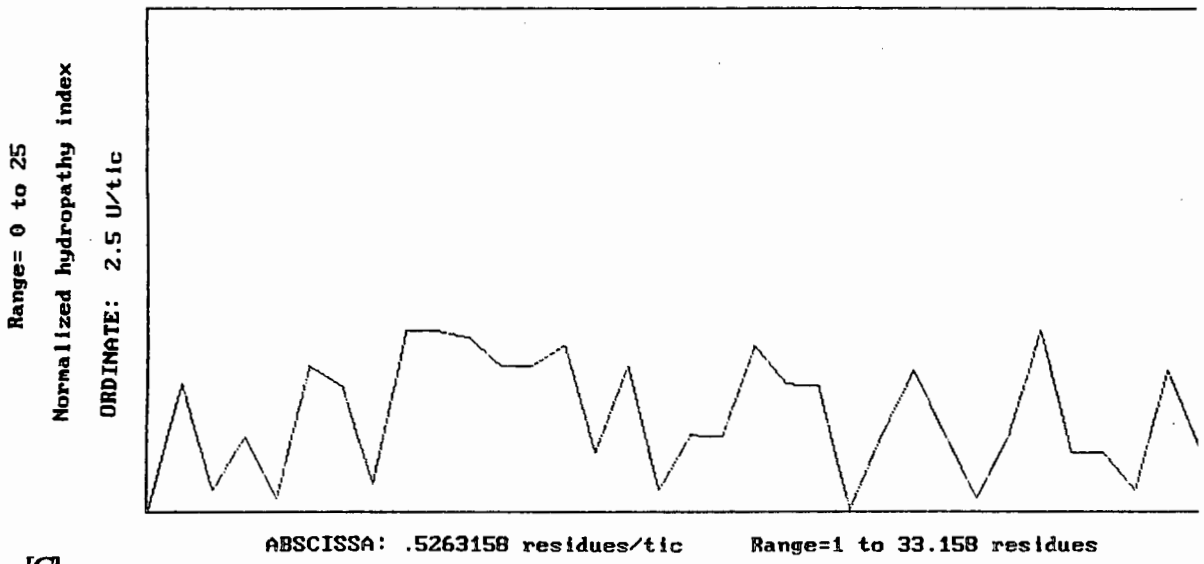


Fig. 4.4 Kyte-Doolittle hydropathy plots of fusion protein Tomwipig-Plant-KY from Pepplot (GCG), (window: 9 residues).

As this window is set over 9 residues for the Kyte-Doolittle plot (GCG), a hydropathy plot with a window of one residue with the SAGA!! package (D.L.Maeder) was done (Fig. 4.5A), using the Kyte-Doolittle data set.. The SAGA!! program indicated the actual hydrophilicities per residue (Fig. 4.5B).

[B]



[C]

Pos	AA	HyPhil	Pos	AA	HyPhil	Pos	AA	HyPhil
1	M	-1.900	12	F	-2.800	23	S	0.800
2	E	3.500	13	L	-3.800	24	C	-2.500
3	S	0.800	14	P	1.600	25	T	0.700
4	K	3.900	15	F	-2.800	26	K	3.900
5	F	-2.800	16	E	3.500	27	S	0.800
6	A	-1.800	17	T	0.700	28	I	-4.500
7	H	3.200	18	S	0.800	29	P	1.600
8	I	-4.500	19	L	-3.800	30	P	1.600
9	I	-4.500	20	M	-1.900	31	Q	3.500
10	V	-4.200	21	A	-1.800	32	C	-2.500
11	F	-2.800	22	R	4.500	33	Y	1.300

Fig. 4.5 Kyte-Doolittle hydropathy plots of fusion protein Tomwipig-Plant-KY from programs:
[B] Hydropathy (SAGA!! ; D.L.Maeder), with a window of one residue.
[C] Hydrophilicities of each residue of the Plant-KY fusion protein (Peptidestructure; GCG).

PeptideStructure (GCG) makes secondary structure predictions of an amino acid sequence (Jameson and Wolf, 1988). Hydrophilicity is calculated according to the algorithm of Hopp and Woods (1981). The window, normally set to 7 residues, can be changed with the optional parameter /HWINdow. The Fold program (GCG) predicted an optimal RNA secondary structure, with a predicted stability

indicated by a Gibbs free energy of -9.3, not expected to interfere with translation of the transcript (Fig. 4.6).

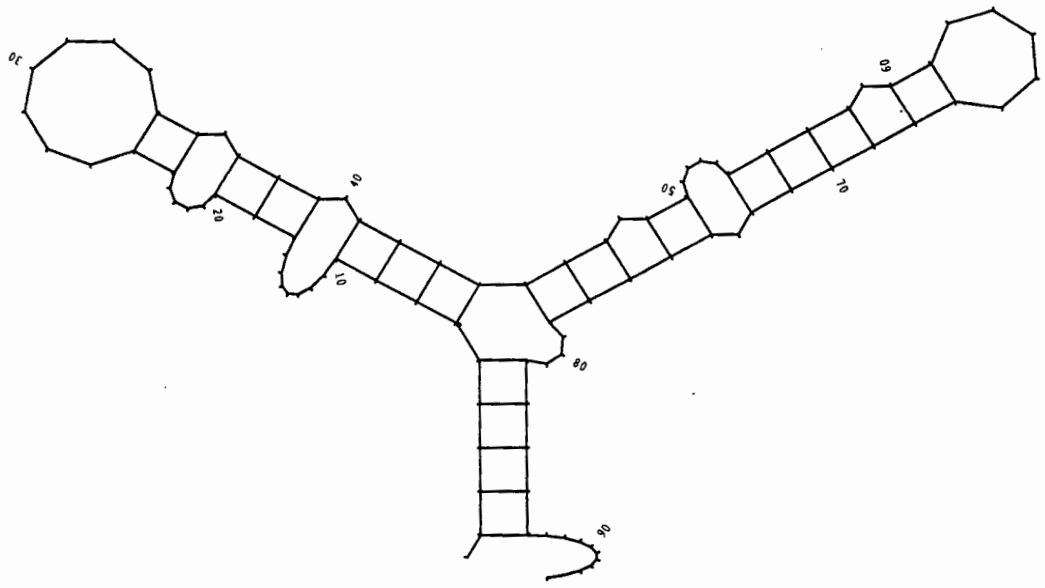


Fig. 4.6 Proposed secondary structure of the mRNA transcript of the Tomwipig-Plant-KY construct and its stability by free energy estimation as predicted by Fold (GCG).

According to the Primer program, the annealing temperature (T_m) of the 5' end of oligo I (75% GC) to the homologous target sequence of the pKS vector was estimated at 52.4°C and that of the 3' end (67% GC), at 57.0°C. These annealing temperatures being close, the ends of oligonucleotide I would be predicted to anneal nearly simultaneously into the cut ends of the pKS vector during the cooling step. This attribute was thus not expected to retard annealing.

4.3.2 Cloning of oligonucleotides by bridge-break repair

For screening of presumptive clones, miniprepared DNA was cut with *Spe*I. A disadvantage of this screening method is that escapee pKS parental molecules would also be cut at the *Spe*I site. Blue/white colour selection was unreliable, as there was erratic colony colour observed including colour streaking. Dot-blot screening of colonies was therfore used after hybridisation with radiolabelled

oligo I probe and a putative clone pKS(TomY23) isolated for further analysis (Fig. 4.7).

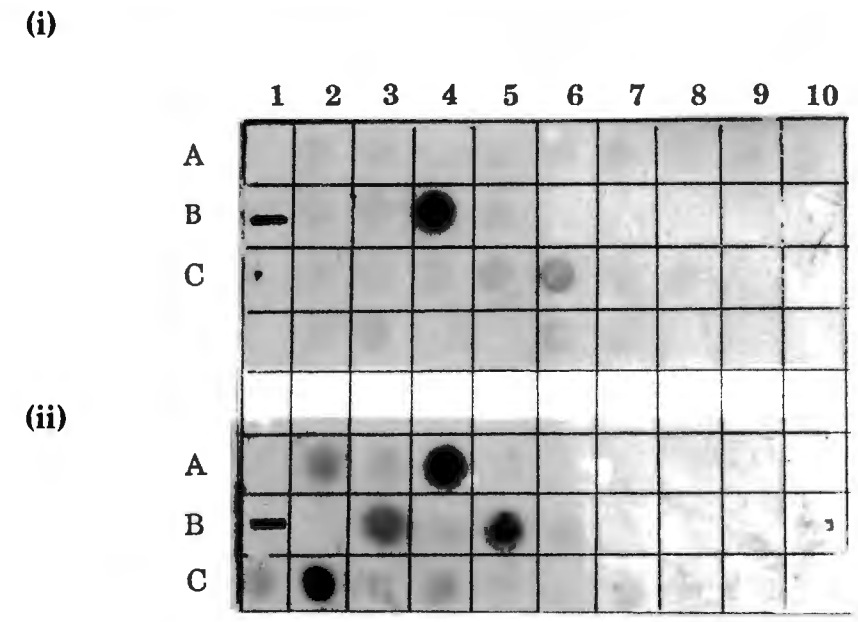


Fig. 4.7 Detection of putative clone pKS(TomY23) by dot-blot hybridisation from two separate blots. (i) and (ii) B1, negative control pKS. (i), B4 and (ii) A4, putative clone pKS(TomY23).

However, upon sequencing, pKSTomY23 was found to be a false positive. As DNA from this colony lit up on a second independent dot-blot (Fig. 4.7.ii), there is a small possibility that oligomer I may have annealled into an ectopic site. In order to investigate this possibility, a FASTA homology search between the pBluescript vector sequence and the full-length oligonucleotide II sequence, including the left and right 15-bp arms was done. This located, in addition to the 15-bp of homology with pKS at positions 684-697, a 10-bp overlap of 100% identity at position 350-360 of pKS (result not shown). On the basis of this result there may be a small chance of the oligonucleotide annealing in here. Otherwise, these

signals could have been false positives due to the 15-bp regions of vector homology within the 5' and 3' flanking regions of the Tomwipig probe.

After a number of cloning parameters had been altered without obtaining recombinants, the oligonucleotide itself was run on a 12% polyacrylamide gel, and was observed to run as a smear (Fig. 4.8A).

This problem was not due to incorrect storage, as material that had been aliquotted and frozen at -20°C immediately after synthesis was found to migrate similarly in acrylamide gels (Fig. 4.8B), as did a subsequently resynthesised preparation (Fig. 4.8C), while the original primers used for double-strand cloning (2.2.3), stored at 4°C for over two years had remained stable (Fig. 4.8B and C, Lanes 1 and 6). Using resynthesised oligo I preparations, (Fig. 4.8B and C), little of the oligonucleotide appeared to be full-length, in which case it was reasoned insufficient viable oligo was being added in each annealing reaction.

The concentration of the full-length oligonucleotide may have been as much as 50 to 100-fold less than indicated by the OD value of the sample, the rest being made up of shorter sequences. Resynthesised oligonucleotide was supplied and a range of oligo concentrations were used to transform cells, keeping the quantity of vector constant.

From annealing reactions with resynthesised oligo I, two positive colonies were detected by dot-blot screening. However, after restriction analysis of the DNA with the enzymes *Apa*I and *Sac*I, these were found not to harbour the sequence of interest. Thus, false positives were being isolated from cloning experiments using dot-blot screening.

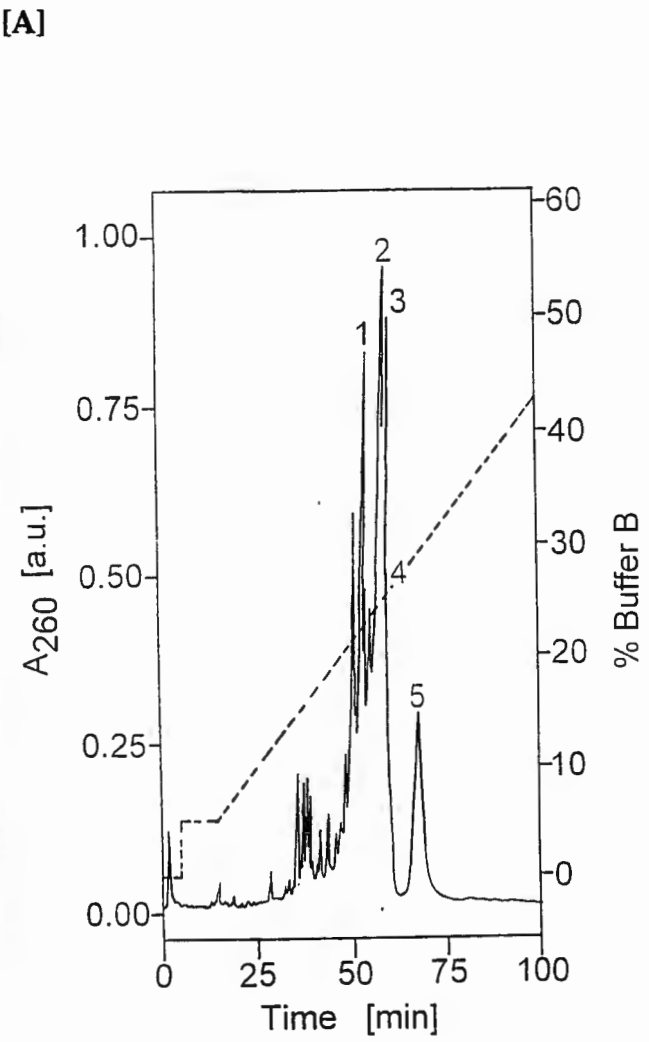


Fig. 4.9 **[A]** HPLC purification profile of oligo 1. **[B]** 12% PAGE analysis of HPLC-purified oligo I. Lanes 1, pBR322/*Hpa*II markers; 2, oligo I before purification; and 3, after purification.

The optimised PCR-screening method for transformants described in section 4.2.3, using two pKS sequencing primers, was found to be more reliable than dot-blot hybridisation in that it produced more quantitative results (Fig. 4.10). PCR reaction products (10 µl) were loaded on an acrylamide gel, on which resolution

was very good, and later on 3% agarose gels (20 µl), on which resolution was sufficient and which could be used more rapidly.

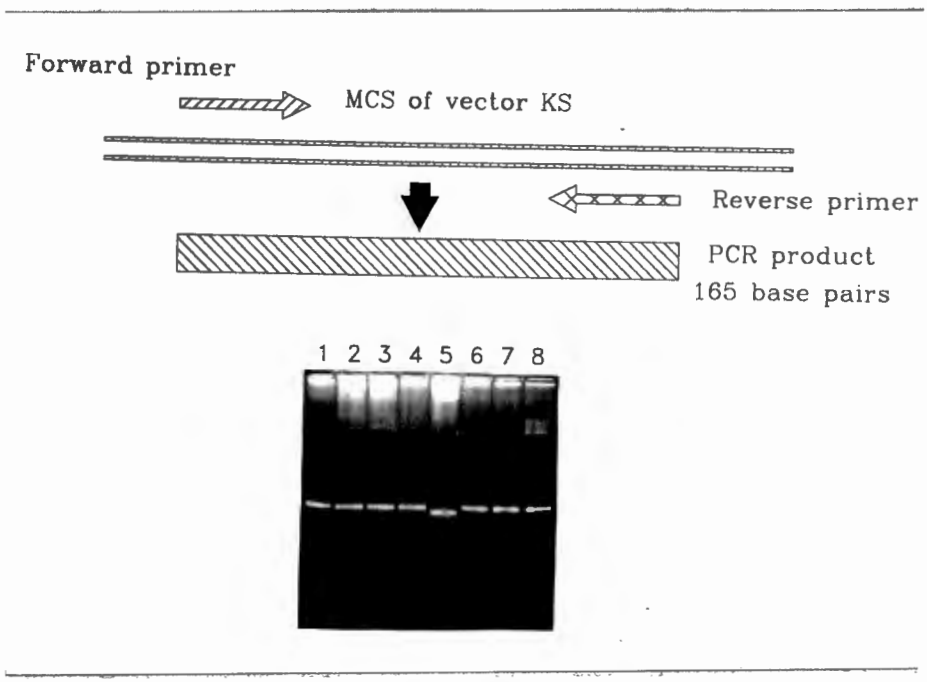


Fig. 4.10 Detection of positive clones in pKS by PCR screening of colonies obtained from annealing experiments with oligo I. Lane 5, negative control: *E. coli* JM105(pKS). Remaining lanes contain positive clones.

Unexpectedly, sequencing of recombinant plasmids identified by PCR screening indicated that oligomer II and not oligomer I had been cloned. Dot-blot hybridisation confirmed the 'oligomer I' supplied was in fact oligo II. After a second synthesis, clones were obtained which showed the same result. The five clones sequenced were all truncated at the 5'-terminus to differing extents. This was probably due to the fact that oligo II has no homology with the vector at this end of the molecule (see Fig. 4.3). Sequence analysis further revealed that rearrangements had occurred in the vector DNA. This could possibly have been due to recombination and deletions caused by the host cell repair machinery during the bridge-break repair and fill-in process due to the lack of vector-complementarity of oligo II at its 5' end. The site 5' to the insert was in three cases

*Hind*III and the downstream site *Apa*I, no *Kpn*I site remaining. The cloning results indicated that experimental conditions were favourable for the bridge-break repair process to occur, which enabled other elements in the experiment to be modified from this point .

Oligo I was re-ordered from Genosys (Cambridge). These suppliers later advised that the oligo would require PAGE purification as it had not been purified, and only 28-30% could be expected to be full-length product, the rest being composed of failure sequences. This was done (Fig. 4.11).

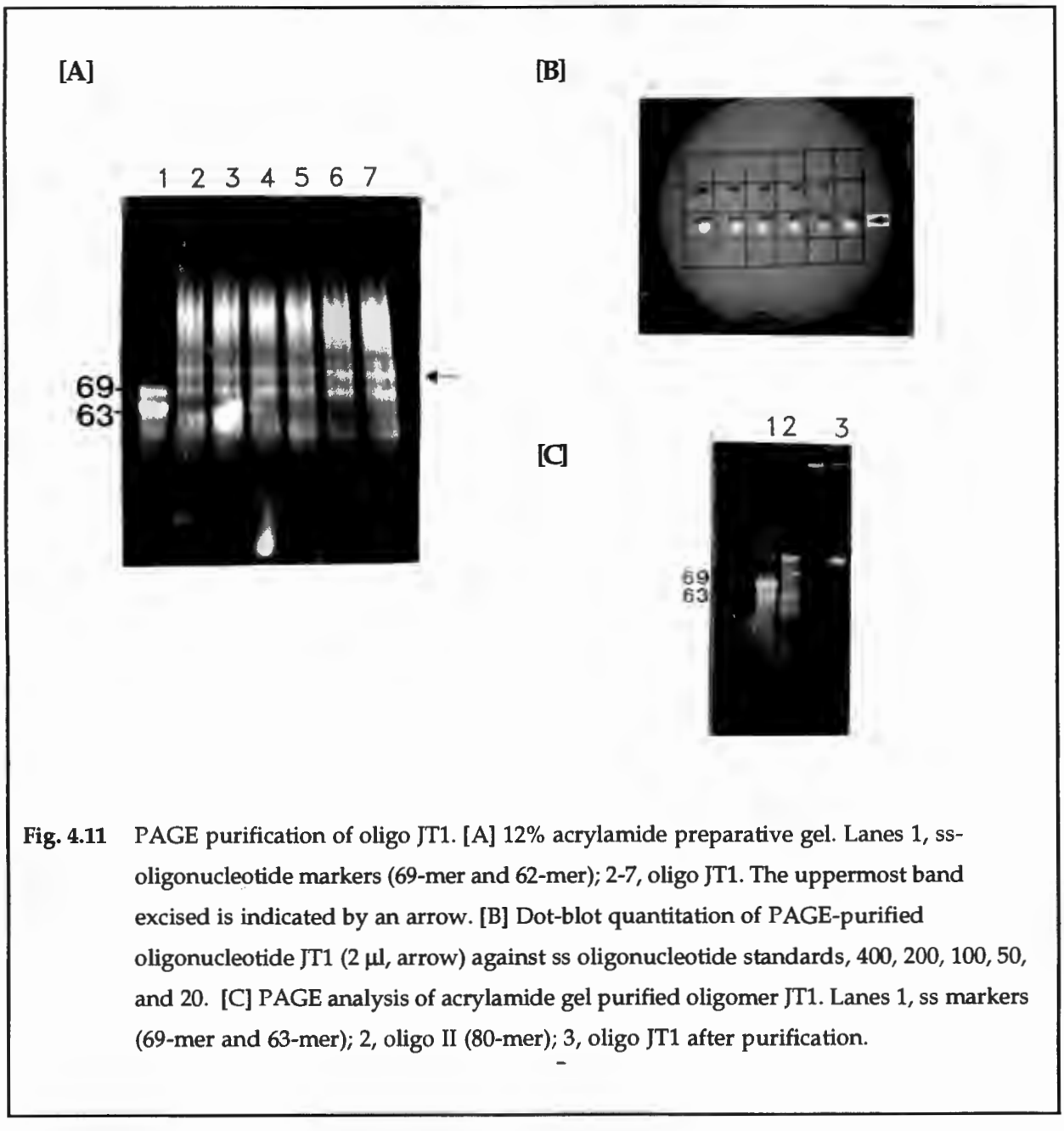


Fig. 4.11 PAGE purification of oligo JT1. [A] 12% acrylamide preparative gel. Lanes 1, ss-oligonucleotide markers (69-mer and 62-mer); 2-7, oligo JT1. The uppermost band excised is indicated by an arrow. [B] Dot-blot quantitation of PAGE-purified oligonucleotide JT1 (2 µl, arrow) against ss oligonucleotide standards, 400, 200, 100, 50, and 20. [C] PAGE analysis of acrylamide gel purified oligomer JT1. Lanes 1, ss markers (69-mer and 63-mer); 2, oligo II (80-mer); 3, oligo JT1 after purification.

The proportion of failure sequences in an oligonucleotide preparation increases with the length of oligonucleotide synthesised, failure sequences becoming a problem for oligonucleotides over 50 bases. Oligo JT1 was therefore electrophoresed through 12% acrylamide preparative gels, the uppermost major band excised, and the oligonucleotide purified from the gel (Fig. 4.11A). After dot-quantitation, preparations of purified oligomer were used in annealing and transformation experiments.

However no transformants were obtained. Investigating this, PAGE-purified 82-mer JT1 was seen to migrate slightly lower than oligo II, an 80-mer (Fig. 4.11C, lane 3), and it was concluded that the predominant molecular species yielding the major band of the original preparation of JT1 had not been of full-length oligomer. This explained the lack of success in obtaining transformants with the purified oligonucleotide. HPLC purification was then chosen to purify a resynthesised stock of oligonucleotide (Fig. 4.12).

Oligonucleotide isolated from HPLC was observed on 10% acrylamide non-denaturing gels to migrate above the 80-mer marker oligo II (Fig. 4.12B). While evidence of some homo-oligomer formation is apparent from these gels, this is expected from synthetic oligonucleotides run on non-denaturing gels. OD readings at A_{260} indicated a total of 570.8 µg of oligonucleotide had been purified, representing a yield of 12% of the DMTr-linked full-length product with respect to the starting material. As an estimated 36% of full-length sequence was expected to have been present in the original sample, this represents a 33% yield for the HPLC purification.

Annealing reactions were done with HPLC-purified oligonucleotide and trouble-shooting the experiment was continued.

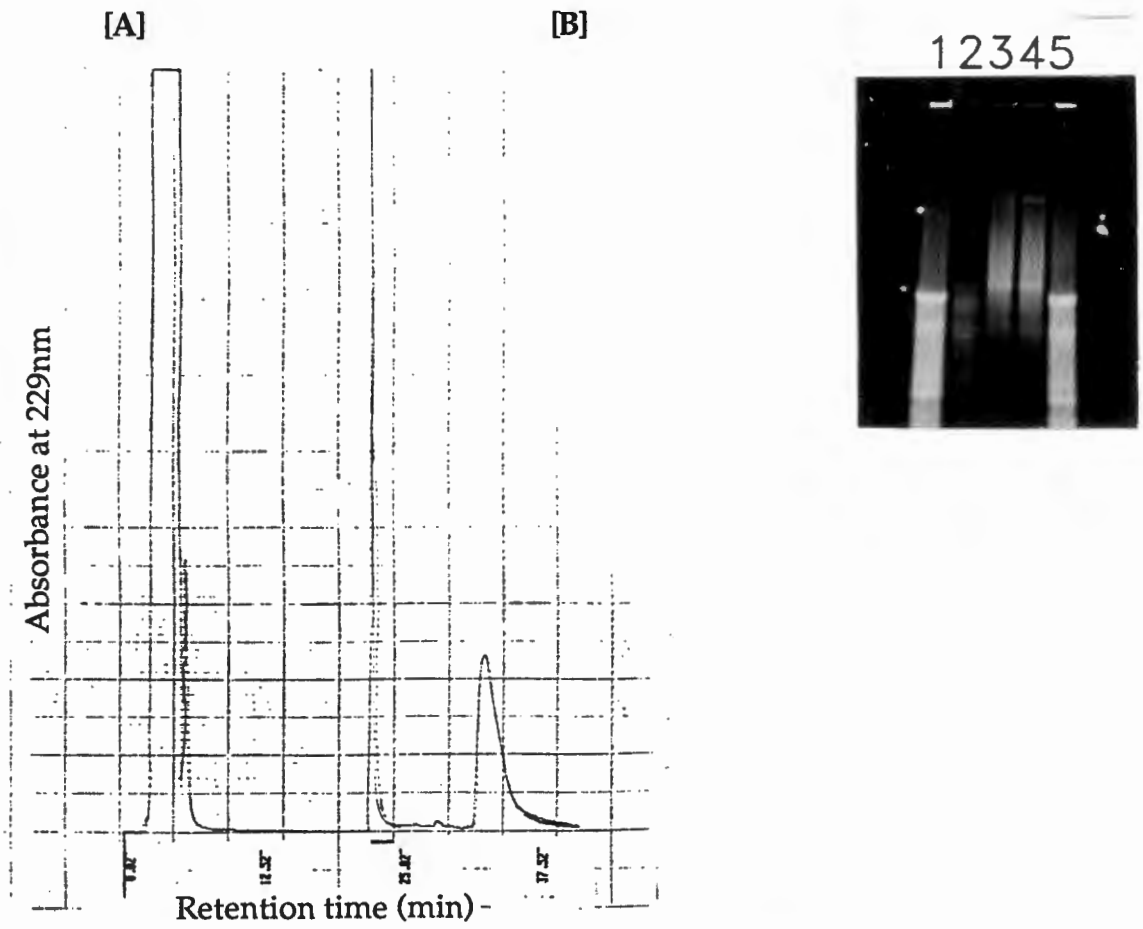


Fig. 4.12 **[A]** HPLC purification profile of oligo JT1 (Genosys). The elution peak is indicated by a small bar under the peak. **[B]** PAGE analysis of HPLC-purified oligo JT1 on a 10% acrylamide non-denaturing gel. Lanes 1 and 5, oligo II (80-mer); 2, oligo JT1 (82-mer) before HPLC purification; 3 and 4, oligo JT1 after HPLC purification.

The temperatures of annealing of each 15-bp arm of the oligo I with the respective vector-homologous sequence were investigated using the Primer program to determine whether a disparity could be reducing the efficiency of annealing. Each 15-bp sequence was entered as a given primer and the respective pKS sequence to which it would anneal entered as the template DNA. Other parameters changed included the following:

1) Reduced vector amounts and addition of more competent cells ie. 200 µl per transformation.

2) Oligonucleotide added at the end of the boiling step in case of degradation over 3 min at 100°C.
3) The secondary structure of Oligo I checked using Fold (GCG) to establish whether this was preventing the annealing of the oligo into the pKS vector. This was not high. A more likely event is homo-dimer formation of the oligonucleotide.
4) Dissociation of possible dimer and oligomer formations of oligos (indicated on non-denaturing gels of Fig. 4.12B) by heating the oligo sample at 100°C for 3 min in a boiling water bath and snap-cooling on ice before incorporation into annealing reactions in 20, 30, 40 and 50 pmol quantities.

From the first transformants obtained from HPLC-purified oligo JT1, PCR-amplified bands on 3% agarose gels migrated slightly higher than those in lanes carrying PCR-amplified pSK (Fig. 4.13B). However upon sequencing two of these plasmids, insert sequence was not found. Upon re-electrophoresis through 3% agarose, the size difference diminished, and may possibly have been due to the greater amount of amplimer produced from the pKS control causing faster migration through the gel matrix (a charge effect). Two further PCR positives were detected, pKS(Tom5) and pKS(Tom6) (Fig. 4.13C).

These appeared from two separate gels to produce amplimers with a size increase expected from inserts of approximately 50-base pairs in the MCS, as compared with those of a previously obtained clone pKS(Tom1), harbouring a 37-bp insert (Fig. 4.13A). After restriction analysis of the maxiprepared DNA with *Kpn*I and *Sac*I on 10% acrylamide, however, these plasmids were found not to harbour an insert. The large size difference observed could not have been an effect of the gel, as the same results were found from two separate gels.

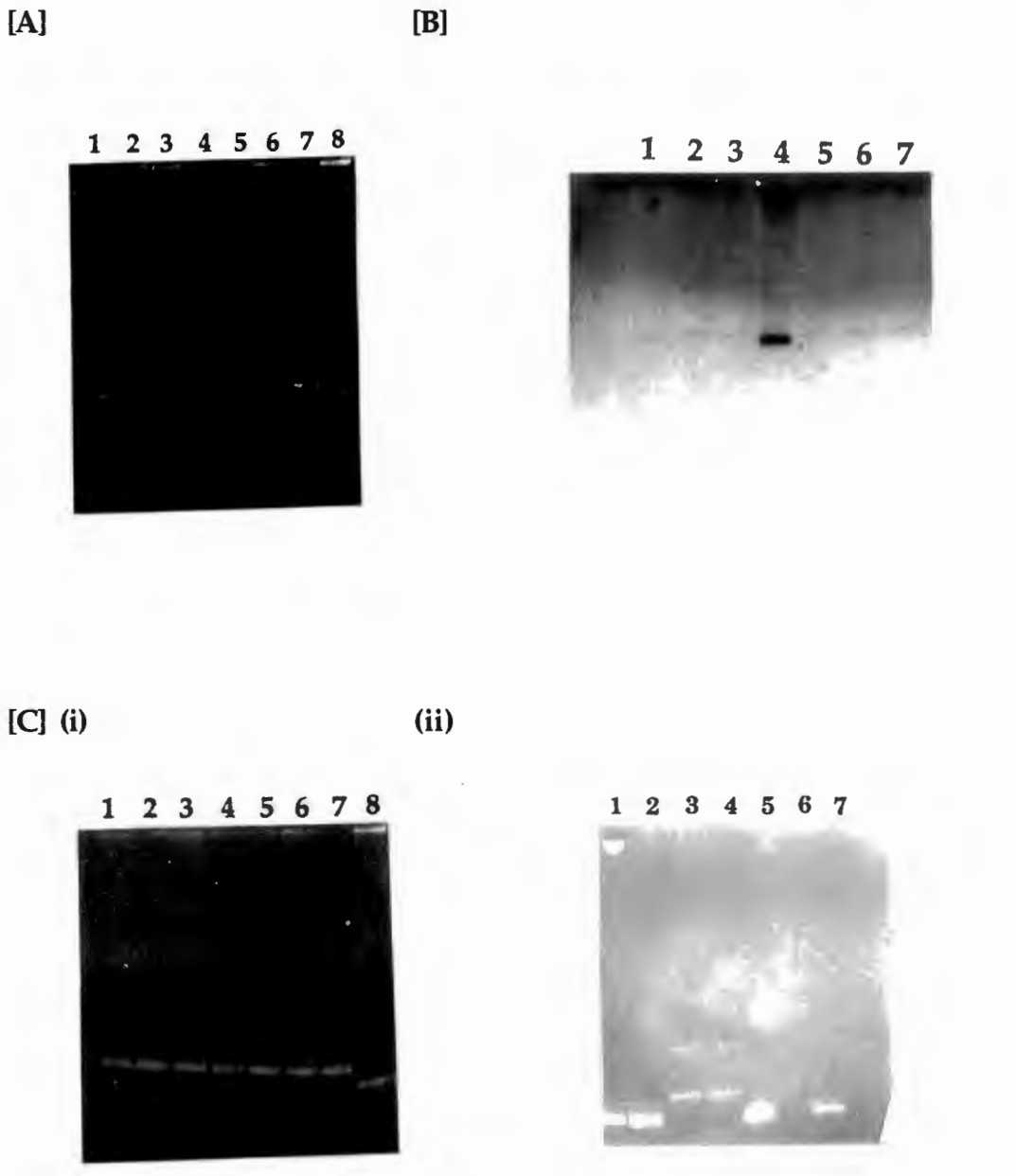


Fig. 4.13 PCR amplimers of presumptive clones of oligomer I in pKS detected by PCR screening on 3% agarose gels. [A] Lanes 1-7, clones pKS(TomI); lane 8, pSK control. [B] Lane 4, pSK control; other lanes, putative clones, including pKS(Tom22), lane 3. [C] (i) Lanes 1-3, negatives; 4, pSK control; 6 and 7, putative clones pKS(Tom5) and pKS(Tom6). [C] (ii) Lanes 1, 2 and 7, negatives; 5, pSK control; 3 and 4, putative clones pKS(Tom5) and pKS(Tom6).

This was an unexpected problem, as no other false positives from PCR screening had been detected thus far with such an obvious size increase. The effect could be due to mis-priming across the MCS of pKS. The problem of the false PCR positives could not be resolved at this time due to time constraints. Usually these could be checked by restriction analysis and discarded if necessary, as they were extremely rare, only being found at this point.

The cloning results of Fig. 4.10 indicated that experimental conditions were favourable for the bridge-break repair process to occur, which enabled other elements in the experiment to be examined and modified. Troubleshooting the annealing experiments was timeconsuming and unfortunately the various unexpected problems with oligonucleotides severely affected the work. Single-strand oligonucleotide cloning can be a rapid method of cloning short synthetic fragments (Mandecki, 1987), but requires high-quality stocks of full-length oligonucleotides with which to perform annealing experiments.

CHAPTER FIVE

THE KY CONSTRUCT FOR DIRECT EXPRESSION IN PLANTS

CONTENTS

CHAPTER FIVE

THE KY CONSTRUCT FOR DIRECT EXPRESSION IN PLANTS

Direct expression of the Plant-KY sequence in plants required this sequence to be subcloned from the pSK-Plant-KY construct (see 2.3.4) into a suitable plant expression vector.

5.1 MATERIALS AND METHODS

5.1.1 Procedures for subcloning KY into pBI121ΔGUS

To excise the Plant-KY sequence, pSK-Plant-KY was cleaved with *Cla*I, single-stranded termini rendered blunt with mung bean nuclease and linearised DNA isolated by TAE gel purification, before digestion with *Bam*HI. The vector pBI121ΔGUS was digested with *Xba*I and recessed 3' termini converted to blunt ends by filling in using the Klenow fragment of *E. coli* DNA polymerase. After heat inactivation of Klenow at 75°C for 10 min the linearised vector was isolated on 0.7% TAE gels and recovered using Geneclean™ (Bio101) followed by digestion with *Bam*HI. Concentrations of DNA fragments recovered with Geneclean™ were estimated by comparison with linearised standard vector DNA samples.

Ligations were carried out at insert:vector ratios of 1:1, 5:1 and 10:1. From calculations of picomolar ends, 4 ng of insert DNA was required (6 pmol) and 874 ng vector pBI121 (6 pmol) for a 1:1 insert to vector ratio in a 20 µl ligation volume. To enable ligation of blunt ends, DNA was ligated at 14°C in a water bath overnight or at 4°C, allowing the water to return to room temperature overnight. A T4 ligase stock found to optimally from test ligations was used at a 10X excess. After transformation of 100 ng DNA by the calcium chloride method, cells were plated on LA plates incorporating kanamycin at 30 µg/ml. Subsequent ligations utilised the 48-mer fragment of pSK-Plant-KY isolated by PAGE, purified from uncut and singly-cut pSK-Plant-KY plasmid, enzymes and excess salts. DNA from the final *Bam*H1 digest of pSK-Plant-KY suspended 1:3 in Orange G tracking dye

was loaded into wells of vertical 12% polyacrylamide gels and electrophoresed at 150 V for $2^1/_2$ h. The DNA co-migrating with the annealed 48-mer Plant-KY control was excised from preparative gels and recovered essentially according to the method of Sambrook (1989), as in 4.2.1.

The concentration of the purified fragment was estimated by comparison with a range of quantities of annealed Bact-KY DNA (54-mer) spotted onto 1% agarose incorporating EtBr (0.5 mg/ml) in a Petri dish. Test DNA was applied in 2 µl aliquots, allowed to dry in a 37°C incubator, and the plate photographed under UV irradiation. The efficiency of blunting the *Xba*I site could not easily be established. Thus, dephosphorylation of the *Xba*I-cut termini of the pBI121ΔGUS vector was employed to eliminate parentals arising from religation of singly-cut vector DNA at this site.

The same ratios of insert: vector DNA used for shotgun ligations were used for ligations of PAGE-purified 48-mer, viz. 1:1, 5:1 and 10:1. In a further set of ligations, DNA was ligated in 100 mM Tris HCl, 5 mM $MgCl_2$, pH 7.6 (Amersham ligation kit) using insert: vector ratios of 60:1 and 600:1. For cloning into pSK, the vector was 60X the size of the insert and a 10:1 insert: vector molar ratio had been successful. For pBI121ΔGUS, the size of the vector was approximately 188X the size of the insert, but proportionally a 31:1 insert: vector ratio could not easily be used (124 ng 48-mer: 874 ng vector) as PAGE-purified 48-mer was available in small quantities. More dilute ligations were therefore done to scale down these quantities, viz. 35 ng insert and 25O ng vector, a ratio of 100:1. To provide for blunt-end ligation, a decreased ATP concentration (0.5 mM), increased T4 ligase (10X), and inclusion of polyethylene gycol (15% PEG 8000, 5 mM MgCl) as a condensing agent, were used. Cells *of E. coli* strains JM105 and LK111 were transformed and plated on LA incorporating kanamycin at 30 µg/ml.

5.1.2. Screening procedures

Colonies obtained from transformations of *E. coli* were screened by dot-blot hybridisation according to the method optimised in 2.3.4, using the Plant-KY oligonucleotide as probe, 5' end-labelled with [γ-^{32}P]dATP with PNK. Blots were prehybridised at 65°C for 1 h and hybridised overnight or for 16 h at 45°C. A PCR screening method was developed for use with the pBI vectors to eliminate false positives obtained from other methods. Although no sequencing primer for GUS was available, a primer with homology within the CaMV 35S promoter of the cointegrate vector pGSJ280 (Deblaere *et al.*, 1987) was obtained. This could be used as a forward primer for PCR, and a reverse primer was required to be designed. The sequence immediately prior to the NOS-ter region, downstream of the site at which GUS had been deleted as a *Sma*I/*Sac*I fragment, was a potential target DNA sequence being a unique region in the plasmid. The NOS-ter of the Agrobacterium Ti plasmid in the pBI plasmids is a 260-bp *Sst*I/*Eco*R1 fragment inserted downstream of GUS (Bevan *et al.*, 1983). The sequence of the junction directly after the *Sac*I site and before NOS-ter was unobtainable from either the suppliers of the pBI vectors (Clontech), a GenEMBL search for pBI plasmids, or from cloning papers. The difficulty of not knowing the NOS-ter-adjacent sequence in the pBI121ΔGUS vector was overcome by sequencing pBI121ΔGUS downstream of the *Sma*I *site* using a forward primer (CaMVFor) with homology within the 3' end of the CaMV 35S promoter (5' GGAAGTTCATTTCATTTGG 3').

To establish where the 150-bp sequence obtained corresponded to the sequence upstream of the NOS-ter, a FASTA (GCG) overlap was done with this sequence and that from a NOS gene sequence located with a GenEMBL search (accession V00087 J01541). The 3' end of the CaMV 35S promoter sequence and the region just prior to the NOS-ter were entered into the Primer program (version 0.5, Whitehead Institute for Biomedical Research) as a joined template DNA sequence of 259 bp, the CaMV 35S promoter sequence sourced from that of the pGSJ280 vector (GenEMBL) and the NOS gene sequence from the pBI121ΔGUS sequence confirmed with NOS-ter sequence from GenEMBL accession V00087

J01541. The CaMV 35S primer was entered as the forward primer to which Primer suggested the reverse primer of the pair.

Experimental parameters optimised for PCR screening pKS constructs were used for pBI121ΔGUS constructs, incorporating the following changes. The annealing temperature was dropped from 50°C to 42°C (experimentally determined and as inferred from the Primer program data- the T_m of the CaMV primer being 48°C). The concentration of the primers 35SFor and ΔRev were checked by PAGE as described (4.2.3). The amount of primer used, of 50 mM stocks, was increased 3-fold as primer concentrations were lower than O.D. values indicated.

As the pBI121 sequencing primer was complementary to the GUS coding strand, a 23-mer sequencing primer was designed with complementarity at the 3' end of CaMV promoter sequence, 5' ATGCTTGTATTTTACCCTATACC 3', designed to be approximately 40 bp from the cloning junction with a GC-rich clamp region at the 3' end over the last 2 bases, and a G+C content over 50%. The primer was checked with the Squiggles Program (GCG) for secondary structure. This primer, however was faulty, and primed up the DNA non-specifically upon sequencing. A primer with homology to the 3' end of the 35S CaMV promoter was later found and used for sequencing.

5.1.3 Agroinfection and plant tissue culture

Mobilisation of the pBI121ΔGUS plasmid and constructs was carried out by the direct Agrobacterium transformation method of Holsters *et al.* (1978). Confirmation of binary vector transformation in Agrobacterium was by minipreparation of the DNA and agarose gel electrophoresis. These DNA samples were further analysed by dot-blot hybridisation analysis using a Plant-KY 5' end-labelled probe as described in 2.2.4. Leaf discs of *Nicotiana tabacum* cv Petit Havana SR1 were transformed by the method of De Block *et al.* (1987). The *A. tumefaciens* strains used were LBA4404(pAL4404) (Hoekma *et al.*, 1983) and C58C1(pMP90) (Koncz and Schell, 1986).

5.2 RESULTS

5.2.1 Subcloning KY into pBI121ΔGUS

i. Shotgun ligations

No signals were obtained upon dot-blot screening of colonies obtained from shotgun ligations and PAGE-purified 48-mer was accordingly used in subsequent ligations.

ii. PAGE isolation of the the 48-mer Plant-KY fragment

Initially, the 48-mer fragment recovered from 17 µg plasmid DNA was not sufficient to be seen on a polyacrylamide gel after PAGE purification. Calculations showed that in order to obtain sufficient 48-mer Plant-KY fragment for ligation reactions, more starting plasmid DNA must be digested. Recovery from ~12 000 ng pSK-Plant-KY could be predicted from calculations as follows:

$$\frac{\text{(Plant-KY)}}{\text{(pSK-Plant-KY)}} = \frac{48\text{ bp}}{(2959\text{ bp} + 48\text{ bp})} = 0.016 \times 12\,000 = 191.5\text{ ng KY}$$

Obtaining an estimated 20% recovery from PAGE purification (P-Y. Ma, pers. comm.) would yield only 38.3 ng of the 48-mer to be used in ligations. Thus batches of 9 tubes of 20 µg each of DNA were prepared at a time for PAGE purification, the product from these being sufficient for estimations of yield by PAGE (Fig. 5.1A and B). Dot-spot detection proved to be an improved method of estimating yield of PAGE-purified DNA, being more sensitive and requiring less product than PAGE analysis (Fig. 5.1C).

The optimised yield of acrylamide-purified Plant-KY ds sequence from the pSK-Plant-KY plasmid clone was estimated at 30 ng/µl, (Fig. 5.1C), representing a total yield of ~450 ng (in ~20 µl) from 7 preparative gels, each of a batch of digests of 20 µg pSK-Plant-KY starting DNA.

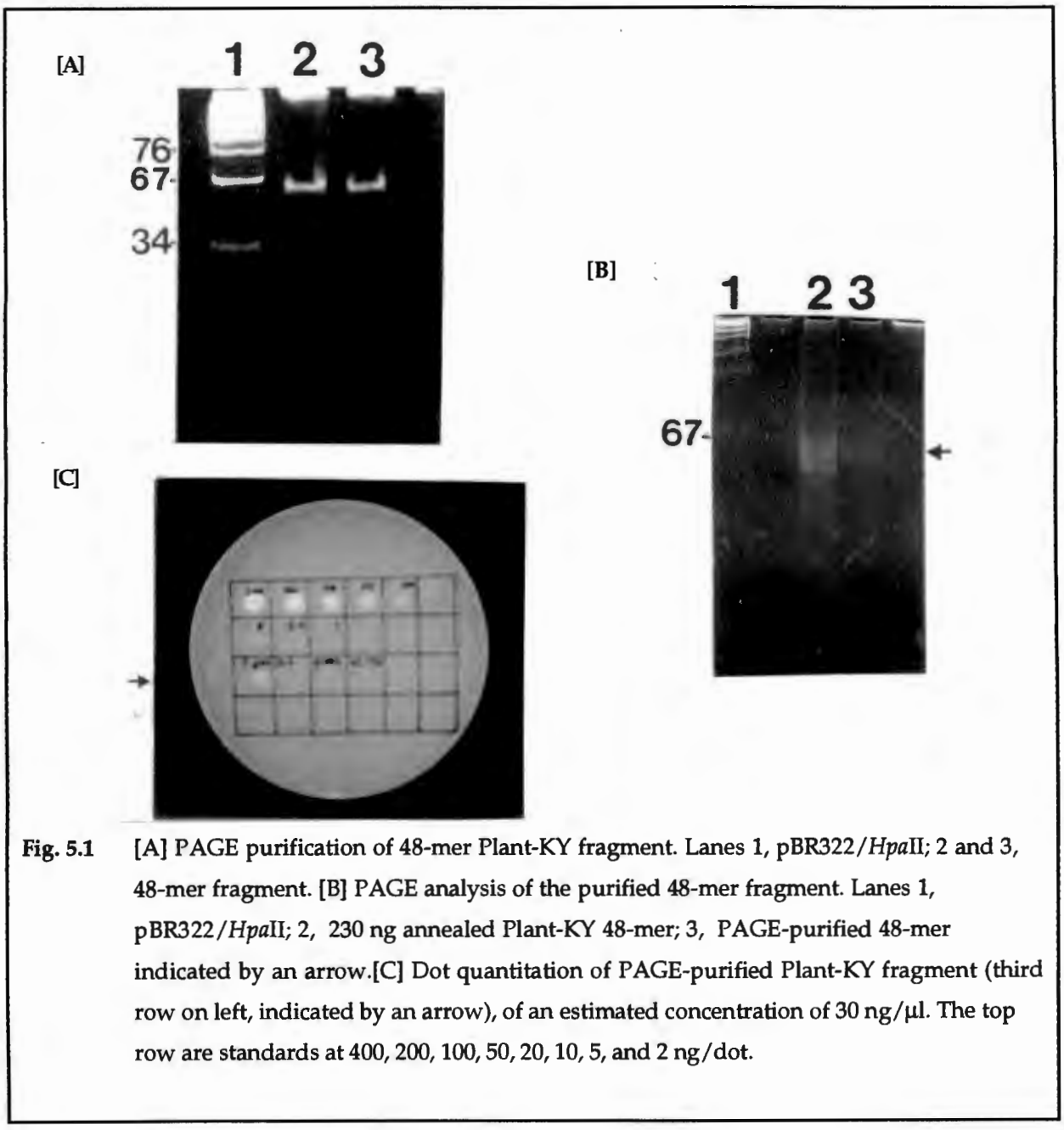


Fig. 5.1 [A] PAGE purification of 48-mer Plant-KY fragment. Lanes 1, pBR322/*Hpa*II; 2 and 3, 48-mer fragment. [B] PAGE analysis of the purified 48-mer fragment. Lanes 1, pBR322/*Hpa*II; 2, 230 ng annealed Plant-KY 48-mer; 3, PAGE-purified 48-mer indicated by an arrow.[C] Dot quantitation of PAGE-purified Plant-KY fragment (third row on left, indicated by an arrow), of an estimated concentration of 30 ng/µl. The top row are standards at 400, 200, 100, 50, 20, 10, 5, and 2 ng/dot.

As no transformants were obtained using dot-blot screening, it was possible that MBN was nibbling back too far, or affecting the *Cla*I site. After this Klenow treatment of ends was used to fill in *Cla*I (pSK-Plant-KY) and *Xba*I (pBI121ΔGUS) 3' recessed ends. A putative clone designated pBI121ΔGUS(Y43) was detected as a signal after dot-blot hybridisation analysis (Fig. 5.2A) and transformed into JM105 and JM83 (RecA$^-$).

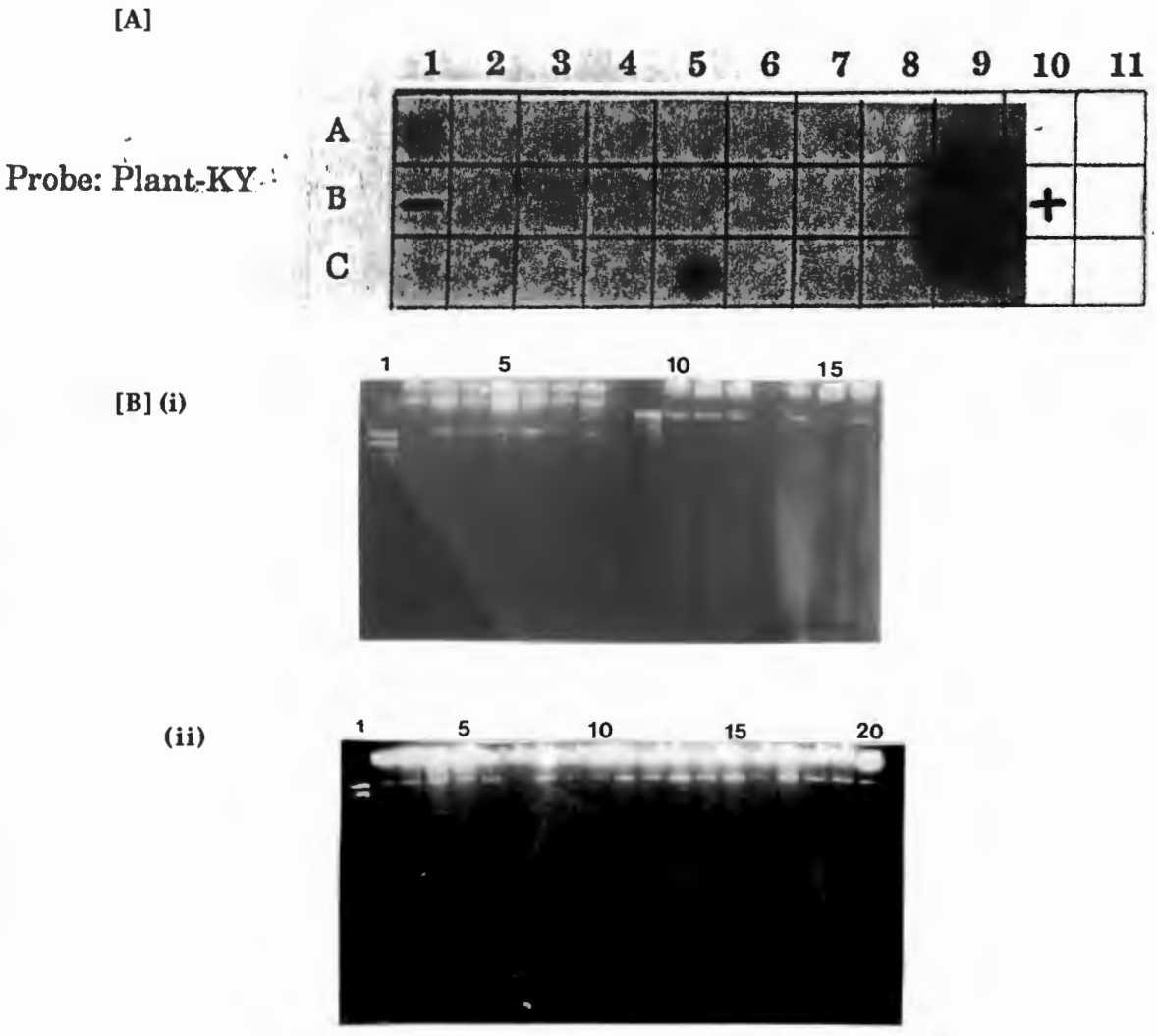


Fig. 5.2 **[A]** Dot-blot hybridisation screening of putative pBI121ΔGUS-Plant-KY clones in *E. coli*. 1B, negative control 5 ng pBI121ΔGUS; 9B, positive control 50 ng Plant-KY oligomer (5'-3' strand); 5C, putative clone pBI121ΔGUS(Y43). Probe: Plant-KY (3'-5' strand). **[B]** Miniprep DNA obtained from pBI121ΔGUS(Y43) transformants of *A. tumefaciens* strains (i) LBA4404 and (ii) C58C1 after freeze-thaw transformation. Lanes1: λ-*Pst*, final lanes, positive control (pBI121ΔGUS), other lanes, miniprep DNA of transformants.

The CaMV 35S sequencing primer was entered into the Primer program as a forward primer to which the program suggested the reverse primer (Fig. 5.3) of the pair. PCR screening with these two primers would produce an amplimer from the region immediately 5' to NOS-ter in the pBI121ΔGUS plasmid.

the pair. PCR screening with these two primers would produce an amplimer from the region immediately 5' to NOS-ter in the pBI121ΔGUS plasmid.

Fig. 5.3 Primer program output file used in designing a reverse PCR primer for pBI121ΔGUS. Forward and reverse primers are indicated by arrows.

The PCR method designed and optimised for pBI121ΔGUS was found to work well as a screening procedure for the pBI plant vectors. This method, however, revealed that a putative clone detected by dot-blot screening pBI121ΔGUS(R1) did not harbour a significantly larger insert sequence across the cloning junction region (Fig. 5.4).

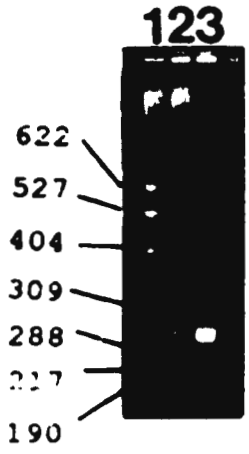


Fig. 5.4 PCR screening of pBI121ΔGUS(R1) Lanes 1, pBR322/*Hpa*II; 2, amplimer of pBI121ΔGUS vector; 3, amplimer of putative clone pBI121ΔGUS(R1).

5.2.3 Agroinfection and plant tissue culture

A high transformation frequency was achieved for the two *A. tumefaciens* strains used, LBA4404 and C58C1 with the plasmids pBI121ΔGUS and pBI121ΔGUS(Y43) as found from minipreparations of DNA from Agrobacterium transformants. All the colonies analysed were found to harbour the plasmid (Fig. 5.2B).

Plasmid DNA from colonies obtained of both strains were miniprepared and analysed by dot-blot hybridisation to confirm binary vector transformations in Agrobacterium. Signals obtained were faint but it was uncertain whether this was due to the low copy number of pBI121ΔGUS or other factors.

After tobacco leaf disc co-cultivation, callus formed at the cut edges of leaf explants after 10-15 days which, on transfer to shoot regeneration medium, developed adventitious shoots and associated callus after $3^1/_2$ weeks. The sections were transferred to rooting medium after 5 weeks. The explants were not further

maintained, however, as sequencing indicated at this stage that pBI121ΔGUS(Y43) did not harbour the 48-mer Plant-KY fragment.

5.3 DISCUSSION

No recombinant clones in pBI121ΔGUS harbouring the Plant-KY coding sequence were found. This could have been due to a problem with one of the many steps in the preparation of vector and/or insert. While the success of a number of these steps were not verifiable, they were checked and methods changed and optimised wherever possible.

It was difficult to accurately assess the amounts of each preparation of 48-mer being added in effect, and what insert: vector ratios to accordingly use, as it was not possible to accurately know the efficiency of Klenow treatment, the efficiency of *Bam*H1 restriction as well as the efficiency of blunt ligation, although test ligations were done to estimate these.

The range of ratios which could be used was restricted as the amount of 48-mer to work with was limiting and required extensive preparation involving a number of handling steps and possible problems could have occurred here.

The number of alternative cloning methods tried in these cloning experiments included two ligation strategies: shotgun ligations and those using PAGE-purified insert DNA. In shotgun cloning, competing ligation events would occur from various competing molecules apart from transformants obtained from uncut pSK-Plant-KY i.e. parentals. Plasmid pSK-Plant-KY singly-cut at the *Cla*I or *Bam*HI site could religate at to produce parentals. Alternatively, these molecules could mop up 48-mer (and any 48-mer not blunted). If after blunting the *Xba*I site of the vector, these molecules did not cut with *Bam*HI, blunt religations at the *Xba*I sites would produce parentals as well and if the vector were not efficiently blunted, vector would religate at adhesive ends of the *Xba*I site. Even if every step proceeded correctly, ligation of the 48-mer back into pSK could be expected to be

favoured over ligation into the pBI121ΔGUS vector due to the smaller size of the former plasmid (2.9 kb vs 9 kb). It would be expected to have been advantageous to have used dilute ligations, due to the relatively large size of the vector.

The advantage of the second ligation method was that most of the above competing ligation events would be eliminated. Disadvantages were the low recovery of the 48-mer upon PAGE purification, and possible damage to the ends of the fragment during purification, although this would be expected to have been compensated for by the lowered number of false transformants from competing ligation events and from uncut parental pSK-Plant-KY molecules. It was found that parental colonies obtained were greatly reduced. However, even for PAGE purified 48-mer, the *Bam*H1 cohesive end ligation of the 48-mer leading to Plant-KY dimer formation, is probably a more favoured event than the ligation of the *Bam*HI site into the vector together with blunt-end ligation with the vector, considering differences in scale, the length of the linearised vector being 188-fold greater than that of the KY insert. Attempts were made to account for this by increasing the amount of the 48-mer Plant-KY sequence, but without success.

A further problem in cloning the 48-mer fragment was that pBI121ΔGUS vector molecules singly cut with *Xba*I could not be distinguished or purified from those additionally cut with *Bam*HI at the second site due to the close proximity of the two sites for *Xba*I and *Bam*HI. The *Bam*HI restriction enzyme requires a 10 bp stretch to cut at an efficiency of 90% to completion and cuts at an efficiency of 25% if only an 8 bp stretch of DNA is available at the site. After cleavage with *Xba*I at the pBI121 cloning junction, fill-in of the 5' overhang would supply a 16 bp footprint for *Bam*HI digestion, but if this did not occur there would be only an 8 bp footprint over which *Bam*HI could cut (25% efficiency over 20 h digestion). Therefore vector molecules not cut by *Bam*HI would recircularise at the blunted *Xba*I site, producing parental vector molecules. While treatment of vector termini with CIP effectively reduced parentals, failure of *Bam*HI to cut would have prevented ligation of the 48-mer into the vector.

As cloning of this construct without a leader would have mainly served as a negative control for the construct with a signal sequence of 4.1.2, it was not further pursued. Not having a signal sequence would preclude the peptide from entering the endoplasmic reticulum and also therefore prevent disulphide bridge formation, necessary for the activity of the peptide. It is doubtful whether the free peptide in the cytoplasm would have been stable, as small foreign peptides in the cytoplasm are most often recognised as foreign and degraded by the proteolytic enzymes of the host plant cell. The PCR screening protocol developed for the pBIDGUS plasmid would be of further use as a rapid and semi-quantitative technique suited for screening transformed plant tissue for the Plant-KY fragment (Langridge *et al.*, 1991; Guidet, 1994) and potentially for other small DNA constructs transferred to plants utilising the pBI vectors.

available, such as the pMal system (Ausubel, 1989). This and another expression strategy (based on metal chelate affinity chromatography) were planned and primers designed but these strategies could not be pursued. It is possible that these expression systems could prove viable for KY peptide production but they may again suffer from the same problems as described for the pET-12a construct (3.4). The lethality observed may be due to the activity of the peptide, interacting with host cell proteases, or due to disruption of the cell membrane during transport across it.

In order to enable the application of KY as a biocontrol agent expressed in plants, two major routes were followed. These were, firstly, the cloning of a peptide sequence comprised of a fusion of the leader sequence of the tomato wound inducible proteinase inhibitor gene (tomwipig) with the Plant-KY sequence (chapter 4) and secondly, the subcloning of the Plant-KY sequence from pSK(Plant-KY) (of chapter 2) into a plant vector. In work towards the former construct, there were a number of problems experienced with the oligonucleotides supplied. Initially, the work towards optimisation of the single-stranded cloning procedure obscured the problems with the oligonucleotides supplied. Upon examining these by PAGE, it was apparent that a range of fragment sizes was present in the preparation of oligonucleotide 1. Further oligonucleotide was supplied twice which had the same problem. Attempts to clone from these preparations by increasing the amount of oligo used in the annealling reactions were not successful, even across a wide range of ratios of oligonucleotide: vector. This problem was apparently overcome when oligo cleanup was done and entire oligo was supplied of the correct molecular weight as observed by PAGE. However, subsequently, oligonucleotide II was twice found to have been supplied instead of oligonucleotide I, upon sequencing clones obtained from these preps. Lastly, using purified preps from a different supplier, a screening problem arose- viz. the appearance of artefactual PCR positive bands using the PCR screening procedure arrived at for selecting the clones. This type of problem is not without precedent in PCR screening applications and may arise from mis-priming. The stringency of the PCR reaction could be increased to prevent this- i.e., increasing the annealing temperature to fine-tune the specificity of amplification. Time constraints had become a problem at this stage, but this

method to obtain an expression construct for heterologous epression of KY in tobacco should prove successful if completed. A large number of annealling experiments were done and troubleshooting was usually being attempted at the mid- to latter stages of the process.

The construct of chapter 5 was difficult to obtain partly due to the small size of the fragment to be subcloned as well as the other problems mentioned. Much troubleshooting was done for these experiments, test ligations being set up to test the efficacy of each step in the cloning, but the exact problem could not be found. This can often be a problem with lengthly cloning procedures in which a number of blind moves are performed, many of which cannot be checked for efficiency due to the decreasing supply of both vector and cloning material.

For possible future work, two other bacterial expression systems could be tried to establish whether the KY inhibitor itself causes lethality or if lethality is occuring due to transport of KY through the membrane. A periplasmic construct could be designed (using pMal for example) and its expression compared with a control construct without leader sequence, a subclone of pSK(Bact-KY) and/or pSK(Bact-KY-Tuft). For the latter, antibodies raised against the KY-Tuft peptide could be used to test for expression of a Plant-KY and /Plant-KY-Tuft peptide expressed in plants.

The design for the expression construct in plants with the tomwipig leader could be further pursued. For this to be effective, the oligonucleotides must be well purified, ideally by PAGE purification. PCR-based screening for the correct constructs followed by sequencing should yield the correct construct in pKS for subsequent directional subcloning into pBI121ΔGUS. Plants rendered transgenic by Agroinfection (or other transformation method) using the resulting plasmid could be screened using the PCR screening method described in 5.1.2ii. Bioassays on regenerated transgenic plants would establish the efficacy of the KY inhibitor. Subsequently, F1 plants could be screened to test for the stable Mendelian inheritance and segregation of the transfered gene. Finally, field resistance could be tested for in field trials using these transgenic plants.

REFERENCES

AN G., MITRA A., CHOI H.K., COSTA M.A., AN K., THORNBURG R.W. and RYAN C.A. (1989) Functional analysis of the 3' control region of the potato wound-inducible proteinase inhibitor II gene. *The Plant Cell* **1**: 115-122

ARKOWITZ R.A., JOLY J.C. AND WICKNER W. (1993) Translocation can drive the unfolding of a preprotein domain. *EMBO* **12** (1): 243-253

ATKINSON A.H., HEATH R.L., SIMPSON R.J., CLARK A.E. AND ANDERSON M.A. (1993) Proteinase inhibitors in *Nicotiana alata* stigmas are derived from a precursor protein which is processed into five homologous inhibitors. *Plant Cell* **5**: 203-213

AUSUBEL F.M., BRENT R., KINGSTON R.E., MOORE D.D. SEIDMAN J.G., SMITH J.A. AND STRUHL J. (1987) Current Protocols in Molecular Biology. John Wiley and Sons, New York

BARBERO J.L., BUESA J.M., PENALVA M.A., PEREZARA A. AND GARCIA J.L. (1986) Secretion of mature human interferon alpha-2 into the periplasmic space of *Escherichia coli. J. Biotech.* **4**(5): 255-267

BEDNAREK S.Y. AND RAIKHEL N.V. (1992) Intracellular trafficking of secretory proteins. *Plant Mol. Biol.* **20**: 133-150

BEVAN M., BARNES W.M., AND CHILTON M. (1983a) Structure and transcription of the nopaline synthase gene region of T-DNA. *Nucl. Acids Res.* **11** (2): 369-385

BEVAN M. FLAVELL R.B., AND CHILTON M.D. (1983b) A chimaeric antibiotic resistance gene as a selectable marker for plant cell transformation. *Nature* **304**: 184-187

BEVAN M. (1984) Binary *Agrobacterium* vectors for plant transformation. *Nucl. Acids Res.* **12**: 8711-8721

BODE W. AND HUBER R. (1992) Natural protein proteinase inhibitors and their interaction with proteinases. *Eur. J. Bioch.* **204**(2): 433-451

BOLTON G.W., NESTER E.W. AND GORDON M.P. (1986) Plant phenolic compounds induce expression of the *Agrobacterium tumefaciens* loci needed for virulence. *Science* **232**: 983-985

BOND J.S. AND BUTLER P.E. (1987) Intracellular proteases. *Ann. Rev. Biochem.* **56**: 333-364

BRAAKMAN I., HELENIUS J. AND HELENIUS A. (1992) Role of ATP and disulphide bonds during protein folding in the endoplasmic reticulum. *Nature* **356**: 260-262

BREATHNACH R. AND CHAMBON R. (1981) Organisation and expression of eukaryotic split genes-coding for proteins. *Ann. Rev. Bioch.* **50**: 349-383

BROADWAY R.M. AND DUFFEY S.S. (1986) Plant proteinase inhibitors: mechanism of action and effect on the growth and digestive physiology of larval *Heliothis zea* and *Spodoptera exiqua. J. Insect Physiol.* **32** (10): 827-833

BROADWAY R.M., DUFFEY S.S., PEARCE G. AND RYAN C.A. (1986) Plant protease inhibitors: a defence against herbivorous insects? *Entomol. exp. appl.* **41**: 33-38

BROSIUS J. (1984) Toxicity of an overproduced foreign gene product in *Escherichia coli* and its use in plasmid vectors for the selection of transcription terminators. *Gene* **27**: 161-172

BROWN W.C. AND CAMPBELL J.L. (1993) A new cloning vector and expression strategy for genes encoding proteins toxic to *Escherichia coli*. *Gene* **127**: 99-103

BRUNKE K.J. AND MEEUSEN R.L. (1991) Insect control with genetically engineered crops. *Tibtech.* **9**: 197-200

BURGESS E.P.J., MAIN C.A., STEVENS P.S., CHRISTELLER J.T., GATEHOUSE A.M.R. AND LAING W.A. (1994) Effects of protease inhibitor concentration and combinations on the survival, growth and gut enzyme activities of the black field cricket, *Teleogryllus commodus*. *J. Insect Physiol.* **40** (9): 803-811

CARRIER M.J. (1983) High expression of foreign proteins in *Echerichia coli* and its consequences. *Trends Biotechnol.* **1**: 109-113

CHEN H-B., WENG J-M., JIANG K. AND BAO J-S. (1990) A new method for the synthesis of a structural gene. *Nucl. Acids Res.* **8** (4) 871-878

CHRISPEELS M.J., AND RAIKHEL N.V. (1992) Short peptide domains target proteins to plant vacuoles. *Cell* **69**: 613-616

CICCARELLI R.B., LOOMIS L.A., McCOON P.E. AND HOLZSCHU D.L. (1990) Insertional gene synthesis, a novel method of assembling consecutive DNA sequences within specific sites in plasmids. Construction of the HIV-1 *tat* gene. *Nucl. Acids Res.* **18**: 5

CLEVELAND T.E., THORNBERG R.W. AND RYAN C.A. (1987) Molecular characterization of a wound-inducible inhibitor I gene from potato and the processing of its mRNA and protein. *Plant Mol. Biol.* **8**: 199-207

CREIGHTON T.E. AND DARBY N.J. (1989) Functional evolutionary divergence of proteolytic enzymes and their inhibitors. *TIBS* 14 Aug. pp 319-325

DAGAN S., TZEHOVAL E., FRIDKIN M. AND FELDMAN M. (1987) Tuftsin and tuftsin conjugates potentiate immunogenic processes: effects and possible mechanisms. *J. Biol. Response Modifiers* **6**: 625-636

DAS A. (1990) Overproduction of proteins in *Escherichia coli*- vectors, hosts and strategies. *Meth. Enzymol.* **182**: 93-112

DAVIDSON E.H., JACOBS H.T., AND BRITTEN R.J. (1983) Eukaryotic gene expression-very short repeats and coordinate induction of genes *Nature* **301**(5900): 468-470

DAVIES D.D. (1992) Physiological aspects of protein turnover. In: *Encyclopaedia of Plant Physiology*, ed. D. Coulter, B. Parthier, **14A**: 189-228. Springer Verlag: Berlin

DAVIS J.M.D., EGELKROUT E.E., COLEMAN G.D., CHEN T.H.H., DOHERTY H.M., SELVERIAN R.R. AND BOWLES D.J. (1988) The wound response of tomato plants can be inhibited by aspirin and related hydroxy-benzoic acids. *Physiol. and Molec. Plant Pathol.* **33**: 377-384

DAVIS J.M.D., EGELKROUT E.E., COLEMAN G.D., CHEN T.H.H., HAISSIG B.E., RIEMENSCHNEIDER D.E., AND GORDON M.P. (1993) A family of wound-induced genes in *Populus* shares common features with genes encoding vegetative storage proteins. *Plant Mol. Biol.* **23**: 135-143

DEBLAERE R., REYNAERTS A., HOFTE H., HERNALSTEENS J.-P., LEEMANS J., AND VAN MONTAGUE M. (1987) Vectors for cloning in plant cells. In: *Recombinant DNA, part D and E (Meth. Enzymol.)*, R. Wu and L. Grossman (Eds). New York, Academic Press

DENEFLE P. (1987) Chemical synthesis of a gene coding for a human angiogenin, its expression in *Escherichia coli* and conversion of the product into its active form. *Gene* **56**: 61-70

DENG T., NOEL J.P. AND TSAI M-D. (1990) A novel expression vector for high level synthesis and secretion of foreign proteins in *Escherichia coli*: overproduction of bovine pancreatic phospholipase A_2. *Gene* **93**: 229-234

DEVEREAU J., HAEBERI P., AND SMITHIES O. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucl. Acids. Res.* **12**: 387-397

DOREL C., VOELKER T., HERMAN E. AND CHRISPEELS M.J. (1988) Secretion is a default pathway in plant cells. *Plant Physiol.* **86**: 84

DOREL C., VOELKER T.A., HERMAN E.M. AND CHRISPEELS M. (1989) Transport of proteins to the plant vacuole is not by bulk flow through the secretory system and requires positive sorting information. *J. Cell. Biol.* **108**: 327-337

DOWSON DAY M.J., ASHURST J.L., MATHIAS S.F., WATTS J.W., WILSON T.M.A. AND DIXON R.A. (1993) Plant viral leaders influence expression of a reporter gene in tobacco. *Plant Mol. Biol.* **23**: 97-109

EDWARDS P.B., AND WANJURA W.J. (1989) Eucalypt-feeding insects bite off more than they can chew: sabotage of induced defences?. *Oikos* **54**: 246-247

EFIMOV V.A., CHAKHMAKHCHEVA O.G AND OVCHINNIKOV Y.A. (1986) Synthetic DNA fragments as useful tools in genetic and protein engineering. *Biotechnology and Genetic Engineering Reviews* **4**: 78-114

ELPIDINA R. (1991) Localisation of a metalloproteinase and its inhibitor in protein bodies of buckwheat seeds. *Planta* **185**: 46-52

ENFORS S-O. (1992) Control of *in vivo* proteolysis in the production of recombinant proteins. *Tibtech.* **10**: 310-315

ENGELMAN D.M., STEITZ T.A. AND GOLDMAN A. (1986) Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. *Ann.Rev.Biophys. Chem.* **15**: 321-353

FANG R-X., NAGY F., SIVASUBRAMANIAM S. AND CHUA N-H. (1989) Multiple *cis* regulatory elements for maximal expression of the cauliflower mosaic virus 35S promoter in transgenic plants. *The Plant Cell* **1**: 141-150

FERNANDES K.V.S., SABELLI P.A., BARRATT D.H.P., RICHARDSON M., XAVIER-FILHO J. AND SHEWRY P.R. (1993) The resistance of cowpea seeds to bruchid beetles is not related to levels of cysteine proteinase inhibitors. *Plant Mol. Biol.* **23**: 215-219

FINLEY, D. AND CHAU V. (1991) Ubiquitination. *Annu. Rev. Cell Biol.* **7**: 25-69

FIRE A., SAMUELS M. AND SHARP P.A. (1984) Interactions between RNA polymerase-II, factors and template leading to accurate transcription. *J. Biol. Chem.* **259**(4): 2509-2516

FLECKER P. (1987) Chemical synthesis, molecular cloning and expression of a gene coding for a Bowman-Birk-type proteinase inhibitor. *Eur. J. Biochem.* **166**: 151-156

FLORA C.B. (1990) Pesticide risk: making decisions. *Plant Disease* **74** (2): 105-108

FRIESEN J.D. AND AN G. (1990) Expression vehicles used in rDNA technology. *Biotechnol. Adv.* **1**: 205-227

FU L.N., YE R.C., BROWDER L.M. AND JOHNSON R.N. (1991) Translational potentiation of messenger-RNA with secondary structure in *Xenopus*. *Science* **251**(4995): 807-810

GALLIE D.R. AND WALBOT V. (1992) Identification of the motifs within the tobacco mosaic virus 5'-leader responsible for enhancing translation. *Nucl. Acids Res.* **20** (17): 4631-4638

GALLIE D.R. (1993) Posttranscriptional regulation of gene expression in plants *Ann. Rev. Pl. Physiol. Pl. Mol. Biol.* **44**: 77-105

GARCIA P.D., GHRAYEB J., INOUYE M. AND WALTER P. (1987) Wild-type and mutant signal peptides of *Escherichia coli* outer membrane protein interact with equal efficiency with mammalian signal recognition particle. *J. Biol. Chem.* **262**(20): 9463-9468

GATEHOUSE A.M.R., GATEHOUSE J.A., DOBIE P., KILMINSTER A.M. AND BOULTER D. (1979) Biochemical basis of insect resistance in *Vigna unguiculata. J. Sci. Food Agric.* **30**: 948-958

GATEHOUSE A.M.R., GATEHOUSE J.A. AND BOULTER D. (1980) Isolation and characterization of trypsin inhibitors from cowpea (*Vigna unguiculata*). *Phytochemistry* **19**: 751-756

GATEHOUSE A.M.R. AND BOULTER D. (1983) Assesment of the antimetabolic effects of trypsin inhibitors from cowpea (*Vigna unguiculata*) and other legumes on development of the Bruchid beetle *Callosobruchus maculatus. J. Sci. Food Agric.* **34**: 345-350

GATEHOUSE J.A., EVANS I.M., CROY R.R.D. AND BOULTER D. (1986) Differential expression of genes during legume seed development. *Phil. Trans. R. Soc. Lond.* B314, 367-384

GATEHOUSE A.M.R. AND HILDER V. (1993) Nutritional evaluation of the trypsin (EC 3.4.21.4) inhibitor from cowpea (*Vigna unguiculata* Walp.) *British J. Nutrition* **68** (3): 783-791

GEOFFROY P., LEGRAND M. AND FRITIG (1990) Isolation and characterization of a proteinaceous inhibitor of microbial proteinases induced during the hypersensitive reaction of tobacco to tobacco mosaic virus. *Molecular Plant-Microbe Interactions* **3**(5): 327-333

GLICK B.R. AND WHITNEY G.K. (1987) Factors affecting the expression of foreign proteins in *Escherichia coli*. *J. Ind. Micro.* **1**: 277-282

GOLDBERG A.L. (1992) The mechanism and functions of ATP-dependent proteases in bacterial and animal cells. *Eur. J. Biochem.* **203**: 9-23

GOMEZ L. AND CHRISPEELS M.J. (1993) Tonoplast and soluble vacuolar proteins are targetted by different mechanisms. *The Plant Cell* **5**: 1113-1124

GOTTESMAN S. AND MAURIZI M. (1992) Regulation by proteolysis: energy-dependent proteases and their targets. *Microbiol. Reviews* **56** (4): 592-621

GRAHAM J.S., PEARCE G., MERRYWEATHER J., TITANI K., ERICSSON L.H. AND RYAN C.A. (1985) Wound-induced proteinase inhibitors from tomato leaves. The cDNA-deduced primary structure of pre-inhibitor II. *J. Biol. Chem.* **260** (11): 6561-6564

GUIDET F. (1994) A powerful new technique to quickly prepare hundreds of plant extracts for PCR and RAPD analysis. *Nucl. Acids Res.* **22**(9): 1772-1773

HABU Y., SAKATA Y., FUKUSAWA K. AND OHNO T. (1993) Ubiquitous nuclear proteins bind to 5' upstream region of major Kunitz chymotrypsin inhibitor gene in winged bean. *Plant Mol. Biol.* **23**: 1139-1150

HAMMOND R.W., FOARD D.E. AND LARKINS B.A. (1984) Molecular cloning and analysis of a gene coding for the Bowman-Birk protease inhibitor in soybean. *J. Biol. Chem.* **259**(15): 9883-9890

HARA-NISHIMURA I., TAKEUCHI Y., AND NICHIMURA M. (1993) Molecular characterisation of a vacuolar processing enzyme related to a putative cysteine proteinase of *Schistosoma mansoni*. *The Plant Cell* **5**:1651-1659

HAYDEN M.A. AND MANDECKI W. (1988) Gene synthesis by serial cloing of oligonucleotides. *DNA* **7** (8): 571-577

HERSHKO A., AND CIECHANOVER A. (1992) The ubiquitin system for protein degradation. *Annu. Rev. Biochem.***61**: 761-807

HILDER V.A., GATEHOUSE A.M.R., SHEERMAN S.E., BARKER R.F. AND BOULTER D. (1987) A novel mechanism of insect resistance engineered into tobacco. *Nature* **330**: 160-163

HILDER V.A., GATEHOUSE A.M.R. AND BOULTER D. (1989a) Potential for exploiting plant genes to genetically engineer insect resistance, exemplified by the cowpea trypsin inhibitor gene. *Pestic. Sci.* **27**: 165-172

HILDER V.A., BARKER R.F., SAMOUR R.A., GATEHOUSE A.M.R., GATEHOUSE J.A. AND BOULTER D. (1989b) Protein and cDNA sequences of Bowman-Birk protease inhibitors from the cowpea (*Vigna unguiculata* Walp.) *Plant Mol. Biol.* **13**: 701-710

HILL R.E., SHAW P.H., BOYD P.A., BAUMANN H. AND HASTIE N.D. (1984) Plasma protease inhibitors in mouse and man: divergence within the reactive centre regions. *Nature* **311**: 175-177

HILL R.E. AND HASTIE N.D. (1987) Accelerated evolution in the reactive centre regions of serine protease inhibitors. *Nature* **326**: 96-99

HOEKEMA A., VANHAARE M.J., FELLINGE A.J., HOOYKAAS P.J. AND SCHILPER R.A. (1985) Non-oncogenic plant vectors for use in the Agrobacterium binary system. *Plant Mol. Biol.* **5** (2): 85-89

HOLLICK J.B. AND GORDON M.P. (1993) A poplar tree proteinase inhibitor-like gene promoter is responsive to wounding in transgenic tobacco. *Plant Mol. Biol.* **22**: 561-572

HOLMGREN A. (1979) Reduction of disulphides by thioredoxin. *J.Biol.Chem.* **254**: 9113-9119

HOLSTERS M., DE WAELE D., DEPICKER A., MESSENS E., VAN MONTAGU M., AND SCHELL J. (1978) Transfection and transformation of *Agrobacterium tumefaciens*. *Mol. Gen. Genet.* **163**: 181-187

HOOYKAAS P.J.J. (1989) Transformation of plant cells via *Agrobacterium*. *Plant Mol. Biol.* **13**: 327-336

HOPP T.P. AND WOODS K.R. (1981) Prediction of protein antigenic determinants from amino acid sequences. *Proc. Natl. Acad. Sci. USA* **78** (6): 3824-3828

HSIUNG H.M. (1986) High level expression, efficient secretion and folding of human growth hormone in *Escherichia coli*. *Biotechnol.* **4** (11): 991-995

HWANG C., SINSKEY A.J., AND LODISH H.F. (1992) Oxidise redox state of glutathione in the endoplasmic reticulum. *Science* **257**: 1496-1502

ICHIHARA S., BEPPU N., MIZUSHIM S. (1984) Protease IV, a cytoplasmic membrane protein of *Escherichia coli* has signal peptide peptidase activity. *J. Biol. Chem.* **259**(15): 9853-9857

INOUYE M., ARNHEIM N. AND STERNGLANZ R. (1973) Bacteriophage T7 lysozyme is an N-acetyl-muramyl-L-alanine amidase. *J. Biol. Chem* **248**: 7247-7252

IVANOV I., CHRISTOV C. AND ALEXCIEV K. (1990) Two methods for rapid assembly and oligermerization of synthetic genes: construction of human calcitonin-encoding sequences. *Gene* **95**: 295-299

JAMESON B.A. AND WOLF H. (1988) The antigenic index: a novel algorithm for predicting antigenic determinants. *Cabios* **4**(1): 181-186

JEFFERSON R.A., BURGESS S., AND HIRSCH D. (1986) β-glucuronidase from *Escherichia coli* as a gene-fusion marker. *Proc. Natl. Acad. Sci. USA* **83**: 8447-8451

JOHNSON K.D., HERMAN E.M., AND CHRISPEELS M.J. (1989) An abundant, highly conserved tonoplast protein in seeds. *Pl. Physiol.* **91**: 1006-1013

JOHNSON A.E., SHENK M.A. AND STEELE R.E. (1993) Protein translocation across the ER membrane: a fluorescent light at the end of the tunnel. *TIBS* **18**: 456-458

JOUBERT F.J., KRUGER H., TOWNSHEND G.S. AND BOTES D.P. (1979) Purification, some properties and the complete primary structures of two protease inhibitors (DE-3 and DE-4) from *Macrotyloma axillare* seed. *Eur. J. Biochem.* **97**: 85-91

JOUDRIER P.E., FOARD D.E. AND LARKINS B.A. (1987) Isolation and sequence of cDNA encoding the soybean protease inhibitors PI IV and C-II. *Plant Mol. Biol.* **10**: 35-42

JUNIPER B.E., ROBINS R.J., AND JOEL D.M., eds. (1989) *The Carnivorous Plants.* **p 353,** Academic Press, London

KEETCH D.P. (1990a) Agricultural entomology in South Africa- yesterday, today and tomorrow. *Plant Protection News.* **21**: 1-2

KEETCH D.P.(1990b) The role of plant protection research in environmental conservation. *Plant Protection News.* **20**: 1-8

KEIL M., SANCHEZ-SERRANO J.J. AND WILLMITZER L. (1989) Both wound-inducble and tuber-specific expression are mediated by the promoter of a single member of the potato proteinase inhibitor II gene family. *EMBO* **8**(5): 1323-1330

KHORANA H.G. (1976) Total synthesis of a gene. *Science* **203**: 614-625

KOCHLAND D. AND BOTSTEIN D. (1980) Secretion of beta-lactamase requires the carboxy end of the protein. *Cell* **20**: 749-760

KOHNO T., CARMICHAEL D.F., SOMMER A. AND THOMPSON R.C. (1990) Refolding of recombinant proteins. *Meth. Enzymol.* **185**: 187-195

KOIJIMA S., OBATA S., KUMAGAI I., AND MIURA K-I. (1990) Alteration of the specificity of the *Streptomyces* subtilisin inhibitor by gene engineering. *Biotechnology* **8**: 449-452

KONCZ C. AND SCHELL J. (1986) The promoter of T_L-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of Agrobacterium binary vector. *Mol. Gen. Genet.* **204**: 383- 396

KONIGSBERG W. AND GODSON G.N. (1983) Evidence for use of rare codons in the *dnaG* gene and other regulatory genes of *Escherichia coli. Proc. Natl. Acad. Sci. USA* **80**: 687-691

KOZAC, M. (1983) Comparison of initiation of protein synthesis in prokaryotes, eukaryotes and organelles. *Microbiol. Rev.* **47**(1): 1- 45

KOZAC M. (1984) Point mutations close to the AUG initiator codon affect the efficiency of translation of rat preproinsulin *in vivo. Nature* **308** (5956): 242-246

KOZAC M. (1986) Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proc. Natl. Acad. Sci. USA* **83**: 2850-2854

KOZAC M. (1991) Structural features in eukaryotic mRNAs that modulate the initiation of translation. *J.Biol.Chem.* **266**: 19867-19870

KADONAGA J.T., GAUTIER A.E., STRAUS D.R., CHARLES A.D., EDGE M.D., AND KNOWLES J.R. (1984) The role of the beta-lactamase signal sequence in the secretion of proteins by *Escherichia coli. J. Biol. Chem.* **259**(4): 2149-2154

KOJIMA M. (1990) Cell surface proteins of *Agrobacterium tumefaciens* involved in the transfer of T-DNA to the host. *Agr. Biol.* **54**(11): 3015-3017

KYTE J. AND DOOLITTLE R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol.Biol.* **157**: 105-132

LANGRIDGE P. LAZZER P. AND LORZ H. (1991) A segment of rye chromosome enhances growth and embryogenesis of calli derived from immature embryos of wheat. *Pl. Cell Rep.* **10**(3): 148-151

LASKOWSKI, Jr. M. AND KATO I. (1980) Protein inhibitors of proteinases. *Ann. Rev. Biochem.* **49**: 593-626

LEEMANS R., REMAUT E.,AND FIERS W. (1989) Correlation between temperature-dependent cytoplasmic solubility and periplasmic export of a heterologous protein in *Escherichia coli. Gene* **85**: 99-108

LITTLE S., CAMPBELL C.J., EVANS I.J., HAYWARD E.C., LILLY R.J. AND ROBINSON M.K. (1989) A short N-proximal region of prochymosin inhibits the secretion of hybrid proteins from *Escherichia . Gene* **83**: 321-329

LAEMMLI U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227** (5259): 680-685

LIEBHABER S.A., CASH F. AND ESHLEMAN S.S. (1992) Translational inhibition by a messenger-RNA encoding region secondary structure is determined by its proximity to the AUG initiation codon. *J. Mol. Biol.* **226**(3): 609-621

LORBERTH C., EBNETH M., AMATI S., S'ANCHEZ-SERRANO J.J. (1993) Promoter elements involved in environmental and developmental control of potato proteinase inhibitor I expression. *Plant J.* **2**: 477-486

LUTKE H.A., CHOW K.C., MICKEL F.S., MOSS K.A., KERN H.F. AND SCHEELE G.A. (1987) Selection of AUG initiation codons differs in plants and animals. *EMBO* **6** (1): 43-48

MADELEY J. (1994) Beyond the pest killers. *New Scientist* 7 May pp. 24-27

MAEDER D.L., SUNDE M. AND BOTES D. (1992) Design and inhibitory properties of synthetic Bowman-Birk loops. *Int. J. Peptide Protein. Res.* **40**: 97-102

MAINA C.V., RIGGS P.D., GRANDEA A.G., SLATKO B.E., MORAN L.S., TAGLIAMONTE J.A., Mc REYNOLDS L.A. AND DI GUAN C. (1988) An *Escherichia coli* vector to express and purify foreign proteins by fusion to and separation from maltose-binding protein. *Gene* **74**: 365-373

MAKOFF A.J., OXER M.D., ROMANOS M.A., FAIRWEATHER N.F. AND BALLANTINE S. (1989) Expression of the tetanus toxin fragment C in *Escherichia coli*: high level expression by removing rare codons. *Nucl. Acids Res.* **17**(24): 10191-10202

MANDECKI, W. (1986) Oligonucleotide-directed double-strand break repair in plasmids of *Escherichia coli*: a method for site-specific mutagenesis. *Proc. Natl. Acad. Sci. USA* **83**: 7177-7181

MANDECKI W. AND BOLLING T.J. (1988) *Fok 1* method of gene synthesis. *Gene* **68**: 101-107

MANLEY J.L. (1983) Analysis of the expression of genes encoding animal messenger RNA by *in vitro* techniques. *Prog. Nucl. A.* **30**: 195-244

MATILE P.H. (1982) Protein degradation. In: *Encyclopaedia of Plant Physiology*, New Ser. **14A**: 169-188, Springer Verlag: Berlin

MATSUI T., SEGALL J., WEIL P.A. AND ROEDER R.G. (1980) Multiple factors required for accurate initiation of transcription by purified RNA polymerase-II. *J. Biol. Chem.* **255** (24): 1992-1996

MAURITZI M.R. (1992) Proteases and protein degradation in *Escherichia coli*. *Experimentia* **48**: 178-201

McBRIDE K.E. AND SUMMERFELT K.R. (1990) Improved binary vectors for Agrobacterium-mediated plant transformation. *Plant Mol. Biol.* **14**: 269-276

McCLAIN W.H., FOSS K., MITTELSTADT K.L., AND SCHNEIDER J. (1986) Variants in clones of gene-machine-synthesised oligodeoxynucleotides. *Nucl. Acids Res.* **14**: 6770

MELLER Y., SESSA G., EYAL Y. AND FLUHR R. (1993) DNA-protein interactions of a *cis*-DNA element essential for ethylene regulation. *Plant Mol. Biol.* **23**: 453-463

MITAGE P.P., WALDEN R. AND DRAPER J. (1988) Vectors for the transformation of plant cells using *Agrobacterium*. In: Plant Genetic Transformation and Gene Expression, A Laboratory manual. Draper, J., Scott R., Armitage P. and Walden R. Blackwell Scientific Publications, New York

MOFFATT B.A. AND STUDIER F.W. (1987) T7 lysozyme inhibits transcription by T7 RNA polymerase. *Cell* **49**: 221-227

MULLET J.E., MURRAY E.E., LOTZER J. AND EBERLE M. (1989) Codon usage in plant genes. *Nucl. Acids Res.* **17** (2): 477-491

MULLET M., KLEIN P.G. AND KLEIN R.R. (1990) Chlorophyll regulates the accumulation of the plastid-encoded chlorophyll apoprotein CP43 and apoprotein D1 by increasing apoprotein stability. *Proc. Natl. Acad. Sci. USA* **87**(11): 4038-4042

NEU H.S. AND HEPPEL L.A. (1965) The release of enzymes from *Escherichia coli* by osmotic shock and during the formation of sphaeroplasts. *J. Biol. Chem.* **240**: 3685-3692

NEUHAUS, J-M., STICHER L., MEINS J.R.F., AND BOLLER T. (1991) A short C-terminal sequence is necessary and sufficient for the targeting of chitinases to the plant vacuole. *Proc. Natl. Acad. Sci. U.S.A.* **88**: 10362-10366

NEURATH H. (1989) Proteolytic processing and physiological regulation. *TIBS* 14 July pp. 268-271

NEURATH H. AND WALSH K.A. (1976) Role of proteolytic enzymes in biological regulation. *Proc. Natl. Acad. Sci. USA* **73** (11): 3825-3835

OAKLEY B.R., KIRSCH D.R. AND MORRIS N.R. (1983) A simplified ultrasensitive silver stain for detecting proteins in polyacrylamide gels. *Anal. Biochem.* **105**: 361-363

ODANI S. AND IKENAKA T. (1978) Studies on soybean trypsin inhibitors. *J. Biochem.* **83**: 737-745

OHSUYE K., NOMURA M., TANAKA S., KUBOTA I., NAKAZATO H., SHINAGAWA H., NAKATA. A AND NOGUCHI T. (1983) Expression of chemically synthesised alpha-neo-endorphin gene fused to *Escherichia coli* alkaline phosphatase. *Nucl. Acids Res.* **11** (5): 1283-1294

PALM C.J., COSTA M.A., AN G. AND RYAN C.A. (1990) Wound-inducible nuclear protein binds DNA fragments that regulate a proteinase inhibitor II gene from potato. *Proc. Natl. Acad. Sci. USA* **87**: 603-607

PARSONS T.J., BRADSHAW, Jr. H.D. AND GORDON M.P. (1989) Systemic accumulation of specific mRNAs in response to wounding in poplar trees. *Proc. Natl. Acad. Sci. USA.* **86**: 7895-7899

PEARCE G., STRYDOM D., JOHNSON S. AND RYAN C.A. (1991) A polypeptide from tomato leaves induces wound-inducible proteinase inhibitor proteins. *Science* **253**: 895-898

PEARSON W.R. AND LIPMAN D.J. (1988) Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* **85** (8): 2444-2448

PENA-CORTES H., SANCHEZ-SERRANO J.J., MERTENS R., WILLMITZER L. AND PRAT S. (1989) Abscisic acid is involved in the wound-induced expression of the proteinase inhibitor II gene in potato and tomato. *Proc. Natl. Acad. Sci. USA* **86**: 9851-9855

PERLAK F.J., FUCHS R.L., DEAN D.A., McPHERSON S.L., AND FISCHHOFF D.A. (1991) Modification of the coding sequence enhances plant expression of insect control protein genes. *Proc. Natl. Acad. Sci. USA* **88**: 3324-3328

PERLMAN D. AND HALVORSON H.O. (1983) A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides. *J. Mol. Biol.* **167**: 391-409

PIECHOCKI M.P. AND HINES R.N. (1994) Oligonucleotide design and optimised protocol for site-directed mutagenesis. *Biotechniques* **16** (4): 702-707

PIERCE J.C. (1994) In-frame cloning of large synthetic genes using moderate-size oligonucleotides. *Biotechniques* **16** (4): 708-715

PLANKOZAC M. (1986) Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proc. Natl. Acad. Sci. USA* **83**: 2850-2856

PUGSLEY A.P. (1993) The complete general and secretory pathway in Gram-negative bacteria. *Microbiol. Reviews* **57** (1): 50-108

REAM W. (1989) *Agrobacterium tumefaciens* and interkingdom genetic exchange. *Ann. R. Phyto.* **27**: 583-618

REAVEY D. AND GASTON K. (1991) Caterpillars in Wonderland. *New Scientist* 23 February pp 52-55

RECHSTEINER M. (1991) Natural substrates of the ubiquitin proteolytic pathway. *Cell* **66**: 615-618

RICHARDSON M. (1977) The plant proteinase inhibitors of plants and microorganisms. *Phytochemistry* **16**:159-169

RINK H., LIERSCH M., SIEBER P. AND MEYER F. (1984) A large fragment approach to DNA synthesis: total synthesis of a gene for the protease inhibitor eglin c from the leech *Hirudo medicinalis* and its expression in *E. coli. Nucl. Acids Res.* **12** (6): 6369-6387

ROSAHI S., ECKES P., SCHELL J., WILLMITZER L. (1986) Organ-specific gene expression in potato: Isolation and characterization of tuber-specific cDNA species. *Mol. Gen. Genet.* **202**: 368-373

RYAN C.A. (1973) Proteolytic enzymes and their inhibitors in plants. *Annu Rev. Pl. Physiol.* **24**: 173-196

RYAN C.A. (1977) Proteinase inhibitors in plant leaves- biochemical model for pest-induced natural plant protection. *Trends bioch.* **3**(7): 148-150

RYAN C.A. (1988) Proteinase inhibitor gene families: tissue specificity and regulation. *In Temporal and Spatial Regulation of Plant Genes.* D.P.S. Verma and Goldberg, eds. (Springer-Verlag: Vienna) pp. 223-236

RYAN C.A. (1989) Proteinase inhibitor gene families: strategies for transformation to improve plant defenses against herbivores. *BioEssays* **10** (1): 20-24

RYAN C.A. (1990) Proteinase inhibitors in plants: genes for improving defences against insects and pathogens. *Ann. Rev. Phytopathol.* **28**: 425-449

RYAN C.A. (1991): Proteinase inhibitors. In: *The Biochemistry of Plants,* Academic Press, New York. pp. 351-370

RYAN C. (1992) The search for the proteinase inhibitor-inducing factor, PIIF. *Plant Mol. Biol.* **19**: 123-133

SAMBROOK J., FRITSCH E.F. AND MANIATIS T. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor

SANCHEZ-SERRANO J.J., KEIL M., O'CONNOR, SCHELL J. AND WILLMITZER L. (1987) Wound-induced expression of a potato proteinase inhibitor II gene in transgenic tobacco plants. *EMBO* **6** (2): 303-306

SANCHEZ-SERRANO J., SCHMIDT R., SCHELL J. AND WILLMITZER L. (1993a) Nucleotide sequence of proteinase inhibitor II encoding cDNA of potato (*Solanum tuberosum*) and its mode of expression. *Mol. Gen. Genet* **203**: 15-20

SANCHEZ-SERRANO J.J., AMATI S., DAMMANN C., EBNETH M., HERBERS K., HILDMAN T., LORBERTH R., PRAT S., AND WILLMITZER L. (1993b) Proteinase inhibitors in the potato response to wounding. In: *Biotechnology in Plant Disease Control.* Wiley-Liss, *Inc. New York* pp. 157-173

SANGER F., NICKLEN S. AND COULSON A.R. (1977) DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**: 5463-5467

SATTAUR O. (1988) A New Crop of Pest Controls. *New Scientist* 14 July: 48-52

SCHECHTER I. AND BERGER A. (1976) *Biochem. Biophys. Res. Commun.* **27**, 157-162

SEBASTIANI F.L., FARRELL L.B., VASQUEZ M., AND BEACHY R.N. (1991) Conserved amino acid sequences among plant proteins sorted to protein bodies and plant vacuoles. Can they play a role in protein sorting? *Eur. J. Biochem.* **199**: 441-450

SHARP P.M., COWE E., HIGGINS D.G., SHIELDS D.C., WOLFE K.H. AND WRIGHT F. (1988) Codon usage in *Escherichia coli, Bacillus subtilis, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Drosophila melanogaster* and *Homo sapiens*: a review of the considerable within-species diversity. *Nucl. Acids Res.* **16** (7): 8207-8211

SHIELDS R. (1987) Towards insect-resistant plants. *Nature* **328**: 12-13

SHUTOV A.D. AND VAINTRAIB I.A. (1987) Degradation of storage proteins in germinating seeds. *Phytochem.* **26**: 1557-1566

SLEAT D.E. AND WILSON T.M.A. (1992) Plant virus genomes as sources of novel functions for genetic manipulations. In: Genetic engineering with plant viruses. T.M.A. Wilson and Davies (eds.) Boca Raton FC. CRC Press. pp. 55-113

SPINDLER K.R. AND BERK A.J. (1984) Translational efficiency of adenovirus early region 1A messenger-RNAs deleted in the 5′ untranslated region. *J. Virol.* **52**(2): 706-710

SPURRIER E.C. (1990) Plant health management issues of public concern: focus on pesticides. Pesticides- there will be change. *Plant Disease* **74** (2): 103-105

STACK J.H., HERMAN P.K., SCHU P.V. AND EMR S.D. (1993) A membrane-associated complex containing the Vps15 protein kinase and the Vps 34 PI 3-kinase is essential for protein sorting to the yeast lysosome-like vacuole *EMBO* **12** (5): 2195- 2204

STADER J.A. AND SILHAVY T.J. (1990) Engineering *Escherichia coli* to secrete heterologous gene products. *Meth. Enzymol.* **185**: 166-187

STANFORD A., BEVAN M. AND NORTHCOTE D. (1989) Differential expression within a family of novel wound-induced genes in potato. *Mol. Gen. Genet.* **215**: 200-208

STORMO G.D., SCHNEIDE T.D. AND GOLD L.M. (1982) Characterisation of translational initiation sites in *Escherichia coli. Nucl. Acids Res.* **10** (9): 2971-2976

STUDIER F.W. AND MOFFATT B.A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high level expression of cloned genes. *J. Mol. Biol.* **189**: 113-130

STUDIER F.W., ROSENBERG A.H., DUNN J.J. AND DUBENDORFF J.W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Meth. Enzymol* **185**: 60-89

STUDIER F.W. (1991) Use of the bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J. Mol. Biol.* **219**: 37-44

STUDIER F.W., ROSENBERG A.H., DUNN J.J. AND DUBENDORFF J.W. (1992) Hosts and vectors for expression in the pET system Novagen technical bulletin

SUMMERS R.G. AND KNOWLES J.R. (1989a) Illicit secretion of a cytoplasmic protein into the periplasm of *Escherichia coli* requires a signal peptide and a portion of the cognate secreted protein. *J. Biol. Chem.* **264** (33): 74-81

SUMMERS R.G., HARRIS C.R. AND KNOWLES J.R. (1989b) A conservative amino acid substitution, arginine for lysine, abolishes export of a hybrid protein in *Escherichia coli*- implications for the mechanisms of protein secretion. *J. Biol. Chem.* **264** (33): 82-88

SUNDE M. (1989) Synthesis of a Bowman-Birk antitrypsin loop and investigation of the structure-function relationships affecting its activity. BSc(Hons.) thesis, Dept. Biochemistry, University of Cape Town

TABOR S. AND RICHARDSON C.C. (1985) A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proc. Natl. Acad. Sci. USA* **82**: 1074-1078

TABOR S. AND RICHARDSON C.C. (1987) DNA sequence analysis with a modified bacteriophage T4 DNA polymerase. *Proc. Natl. Acad. Sci. USA* **84**: 4767-4771

TAKAHARA M., HIBLER D.W., BARR P.J., GERLT J.A. AND INOUYE M. (1985) The *Omp A* signal peptide directed secretion of Staphylococcal nuclease A by *Escherichia coli. J. Biol. Chem.* **260** (5): 2670-2674

TALMADGE K AND GILBERT W. (1989) Cellular location affects protein stability in *Escherichia coli. Proc. Natl. Acad. Sci. USA* **79**: 1830-1833

TAYLOR B.H., YOUNG R.J., AND SCHEURING C.F. (1993) Induction of a proteinase inhibitor II-class gene by auxin in tomato roots. *Plant Mol. Biol.* **23**: 1005-1014

TEIXEIRA A. (1992) Synthesis and expression of the Erythrina trypsin/tissue plasminogen activator (tPA) inhibitor encoding gene. PhD Thesis, University of Cape Town

TERADA S. SATO K., KATO T. AND IZUMIYA N. (1980) Studies on the synthesis of protease inhibitors. *Int. J. Peptide Protein Res.* **15**: 441-454

TOWNSHEND G., BOTES D.P. AND VISSER L. (1982) Properties of fragments prepared from Bowman-Birk inhibitors. *Biochim. Biophys. Acta* **701**: 346-356

VAN RIE J. (1991) Insect control with transgenic plants: resistance proof? *Tibtech.* **9**:177-179

VIERSTRA R.D. (1993) Protein degradation in plants. *Annu. Rev. Plant Physiol. Plant. Mol. Biol* **44**: 385-410

VOGEL (1988) Susceptible regions map to interdomain junctions of protease inhibitors. *J. Biol. Chem.* **263**: 12661-12668

VON HEIJNE G. (1991) Computer analysis of DNA and protein sequences. *Eur.J. Biochem* **199**: 233-236

VON HEIJNE G. AND BLOMBERG C. (1979) Trans-membrane translocation of proteins -the direct transfer model. *Eur. J. Biochem.* **97**: 175-181

WALKER-SIMMONS M. AND RYAN C.A. (1977a) Wound-induced accumulation of trypsin inhibitor activities in plant leaves. Survey of several plant genera. *Plant Physiol.* **59**: 437-439

WALKER-SIMMONS, M. AND RYAN C.A. (1977b) Immunological identification of proteinase inhibitors I and II in isolated tomato leaf vacuoles. *Plant Physiol.* **60**(1): 61-63

WEBB E.C. (1984) *Enzyme nomenclature.* New York: Academic Press

WEN L., HUANG J-K., ZEN K.C., JOHNSON B.H., MUTHUKRISHNAN S., MACKAY V., MANNEY T.R., MANNEY M. AND REECK G.R. (1992) Nucleotide sequence of a cDNA clone that encodes the maize inhibitor of trypsin and activated Hageman factor. *Plant Mol. Biol.* **18**: 813-814

WIECZORCK M., OTLEWSKI J., COOK J., PARKS K., LELUK J., WILIMOWSKA-PELC A., POLANOWSKI A., WILUSZ T. AND LASKOWSKI Jr. M. (1985) The squash family of serine proteinase inhibitors. Amino acid sequences and association equilibrium constants of inhibitors from squash, summer squash, zucchini, and cucumber seeds. *Biochem. Biophys. Res. Commun.* **126** (2): 646-652

WILSON K.A. AND TAN-WILSON A.L. (1987) Characterisation of the proteinase that initiates the degradaton of the trypsin inhibitor in germinating mung beans (*Vigna radiata*). *Plant Physiol.* **84**: 93-98

WILSON K.A., PAPASTOITSIS G., HARTL P., AND TAN-WILSON A. (1988) Survey of the proteolytic activities degrading the Kunitz trypsin inhibitor and glycinin in germinating soybeans (*Glycine max*). *Plant Physiol.* **88** (2): 355-360

WOLFSON J.L. AND MURDOCK L.L. (1987) Suppression of larval Colorado potato beetle growth and development by digestive protease inhibitors. *Entomol. exp. appl.* **44**: 235-240

WONG E.Y., HIRONAKA C.M. AND FISCHOFF D.A. (1992) *Arabidopsis thaliana* small subunit leader and transit peptide enhance the expression of *Bacillus thuringiensis* proteins in transgenic plants. *Plant Mol. Biol.* **20**: 81-93

XU D., MCELROY D., THORNBURG R.W. AND WU.R. (1993) Systemic induction of a potato pin2 promoter by wounding, methyl jasmonate, and abscisic acid in transgenic rice plants. *Plant Mol. Biol.* **22**: 573-588

YAMAGISHI K., MITSUMORI C., TAKAHASHI K., FUJINO K., KODA Y. AND KIKUTA Y. (1993) Jasmonic acid-inducible gene expression of a Kunitz-type proteinase inhibitor in potato tuber discs. *Plant Mol. Biol.* **21**: 539-541

YOO B.Y. AND CHRISPEELS M.J. (1990) The origin of protein bodies in developing soybean cotyledons: a proposal. *Protoplasma* **103**: 201-204

YU L.M., LAMB C.J. AND DIXON R.A. (1993) Purification and biochemical characterisation of proteins which bind to the H-box *cis*-element implicated in transcriptional activation of plant defence genes. *The Plant Journal* **3** (6), 805-816

ZUKER M. AND STIEGLER P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxilliary information. *Nucl. Acids Res.* **9** (1): 133- 148